

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:36 ; Search time 7810.03 Seconds  
(without alignments)  
-7.776 Million cell updates/sec

Title: US-08-956-991-6

Perfect score: 20  
Sequence: 1 cctgattgacctgcaggaag 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_dal.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vl.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_in.\*  
21: em\_om.\*  
22: em\_or.\*  
23: em\_ov.\*  
24: em\_pat.\*  
25: em\_ph.\*  
26: em\_pl.\*  
27: em\_ro.\*  
28: em\_sts.\*  
29: em\_sy.\*  
30: em\_un.\*  
31: em\_vl.\*  
32: gb\_htg1.\*  
33: gb\_htg2.\*  
34: gb\_in1.\*  
35: gb\_in2.\*  
36: em\_ba1.\*  
37: em\_ba2.\*  
38: em\_hum3.\*  
39: em\_hum4.\*  
40: gb\_pr4.\*  
41: gb\_htg3.\*  
42: gb\_htg4.\*  
43: gb\_htg5.\*  
44: gb\_htg6.\*

45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	20	100.0	6110	11 AF023449	AF023449 Homo sapi
2	20	100.0	6413	11 AF023450	AF023450 Homo sapi
3	20	100.0	145861	11 AF064862	AF064862 Homo sapi
4	20	100.0	159424	11 AF064865	AF064865 Homo sapi
5	17.4	87.0	53718	43 AC013576	AC013576 Homo sapi
6	17.4	87.0	89948	11 HS756923	AL035681 Human DNA
7	17.4	87.0	106319	11 HS7477L4	AL009051 Homo sapi
8	17.4	87.0	116967	41 AC009389	AC009389 Drosophila
9	17	85.0	139134	45 AC011140	AC011140 Homo sapi
10	16.8	84.0	3374	9 HSMYOH	Y00821 Human mRNA
11	16.8	84.0	4060	12 MUSA3A	M28383 Mouse A3 m
12	16.8	84.0	4350	12 RN035022	U35022 Rattus norv
13	16.8	84.0	4437	9 HSBMHC	X51592 Human MHC m
14	16.8	84.0	5259	9 HUMDMP	J05211 Human Desmo
15	16.8	84.0	5866	3 AB025262	AB025262 Sus scrof
16	16.8	84.0	5925	40 AF111785	AF111785 Homo sapi
17	16.8	84.0	5929	3 AB025261	AB025261 Sus scrof
18	16.8	84.0	5932	3 AB025260	AB025260 Sus scrof
19	16.8	84.0	5956	1 AF111784	AF111784 Homo sapi
20	16.8	84.0	6010	5 HSMYOSIN	Z38133 H. sapiens m
21	16.8	84.0	6016	11 HUMMYOHP	M36769 Homo sapien
22	16.8	84.0	9588	40 HUMDPI	M77830 Homo sapien
23	16.8	84.0	33593	6 AF063097	AF063097 Bacteriop
24	16.8	84.0	64356	11 HS512B11	AL031058 Human DNA
25	16.8	84.0	104913	11 AC004242	AC004242 Homo sapi
26	16.8	84.0	125669	41 AC011285	AC011285 Homo sapi
27	16.8	84.0	125937	11 HS69M21	AL031735 Human DNA
28	16.8	84.0	198582	11 AC005291	AC005291 Homo sapi
29	16.8	84.0	204082	11 AC001337	AC005333 Homo sapi
30	16.8	84.0	225747	41 AC008083	AC008083 Homo sapi
31	16.4	82.0	758	11 HSLRPS25	AF058421 Homo sapi
32	16.4	82.0	1284	11 AF022150	AF022150 Homo sapi
33	16.4	82.0	1284	11 HSAF000972	AF000972 Homo sapi
34	16.4	82.0	1982	11 AF033021	AF033021 Homo sapi
35	16.4	82.0	2238	11 AF022797	AF022797 Homo sapi
36	16.4	82.0	28690	41 AC009072	AC009072 Homo sapi
37	16.4	82.0	30543	41 AC008618	AC008618 Homo sapi
38	16.4	82.0	42709	41 AC011520	AC011520 Homo sapi
39	16.4	82.0	103194	8 ATAC06223	AC006223 Arabidops
40	16.4	82.0	126956	11 HS8B1	AL053701 Human DNA
41	16.4	82.0	166465	32 HS1447E21	AL050336 Homo sapi
42	16	80.0	2875	12 MUSMHOAMB	M20985 Mouse MHC c
43	16	80.0	2945	12 AF016308	AF016308 Mus muscu
44	16	80.0	45000	43 AC013257	AC013257 Leishmani
45	16	80.0	65386	43 AC015461	AC015461 Homo sapi

#### ALIGNMENTS

RESULT 1  
AF023449 6110 bp mRNA PRI 01-JUN-1998  
LOCUS AF023449  
DEFINITION Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)  
ACCESSION AF023449  
VERSION AF023449.1 GI:3169765

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 6110)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a  
Down Syndrome Region and is Involved in the Development of the  
Nervous System  
Unpublished  
2 (bases 1 to 6110)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
Direct Submission  
Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research  
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los  
Angeles, CA 90048-1869, USA  
Location/Qualifiers  
1. 6110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21 (trisomy 21)"  
/map="21q22, between HMGI4 and MXI"  
/tissue\_type="brain"  
/dev\_stage="14 weeks, fetal"  
/clone="CHD2-42"  
/note="derived from alternately-spliced mRNA"  
1. 6110  
/gene="DSCAM"  
/note="member of immunoglobulin superfamily; involved in  
nervous system development"  
/product="Down syndrome cell adhesion molecule"  
/protein\_id="AAC17966.1"  
/db\_xref="GI:3169766"  
/translation="VESEDLHSLYEYNASLOEVEFASTTGLVPCPAAGIPVTLRM  
VLATGEIYDVGIRVHPNGTLOIPPEPSSSTLIHNTYCTAENSGKRSQDV  
HIRAVIREPTVREDOKTRMGVAVAKCIIPSSVEAFITVVEKDYDTSVSGSFL  
ITSGALYIKDYNEDGLYRCITRRTTGETRQNSATLEPDSANAPSLIDFED  
HRKMAGORVELPCKALGHEPDRLWKDNMPLSLGREFOKVTGLIENIRSDGS  
YCEVSNRGYTAIVIGRLVYKPLKATISPRKXSSGSOVSLSCVSTEDDELSTWY  
RNGELINPGKNRITGIGNHENLIMDHVKSDBGAYOCFYKDKLSADQVVOYVLEDT  
PKIIISAFSEKVVSPAEVSLMCNVKGPPLTITWTLDDPILGSGSHRISOMTSGN  
VSVLNTSSSOVDGQVYRGTANNSAGVLYOARINVRPASIRPKMTITAIAGRTY  
IHCYVIGYPIYSIKYKNSLPLPNROVAFENNGTLKLSQVKEVEDEEITCNVLYO  
POLSTQSVHVTYKVPFTIOPEEPERSIQGVFICVAVSGDPLITTWQDGRIP  
GSLGVTIDNIDFTSLRIISLIMHNGNTICIAARNEAAVEHQSQLIVVPRFVQ  
RDDGIGYKAVILNCSAEGYPTIVYKFSKAGYQFOPALNGRIQVLSGLIK  
HVEEDGYGLCKVNSGADVSKSMILTVKIPAMITSYPTLATQSGKRESCAH  
GKRTIIRWEKEDRIINPEMAYLVSTKEGVEVITITOLIPVREDSGCEKATNS  
YGEDGILQTOPEPDRPELEIKDKVAKRTITARMGPDGNSPTITGYIECKNSDS  
WDSQRTKDYSPQNSKTIIDIPSSITSIRMTAKNRIGKSEPSNELITADBAADG  
PPOEVHLEPISSQIRVTWAPRKLQNGIIRGOYQIGREYSTGNGFONIIISVDG  
DSEVTLIDNLNKTQYGLVQACNRAGTGPSSOEITITLEDYSPSPENVAATSP  
ESISISWTSIKALNGILOGFVVIYMANIMDEGEIKNITTOPLSELDEGKXTN  
YSIOVLAFTRAGDGVSEIOFTRTKEDVGPAGVAKAASASAMVSMPLPKINGI  
IRKTYVCSHPYVUISEPASPSFSYRPNLSRNOVSANVAVNTSGRNSSEII  
TVERPLAKAPRILITSTYTTTMMKQIVLPCKAVGSPVAVMMKSNQSPSLVITDG  
RRIISFSGSFIINTVAEDSGYSCIANNNWSDGDEILNLOVAPPDQPLVLSKTS  
SSITLWLPDNGSSIRGYILIOYSEDSNEMGSPISPSERSYLENLCGTWKT  
LTQNGVGPRISEIIEAKLKEPOFSKOELFASINTRVNLIGNNDGCPITS  
FTLEVPFGTTVTAORTLSKSYLLYDQEAETELQMRVNSAGCEKAKRANT  
NYGSGTIPPLIKSVONEGLTINELGKLMTSCILVGLVLLFVLLLYVARRRBR  
LKLRLRAKSLAELMKNTRTSTLTLMKOOTLBMHIDIPRAOLILEERTWTITDRS  
TVLLIDQDEAKOKSLIVTHVYOSVQATPLVDVSDAPGNPTTRNNAAG  
TANRLASQWTLNRPHTISAHTLTIDWRLPTPRAAGSVKEDSYSPSODTBRAR  
SNVSTESASTYEELARAEHAKMEOLRHAFTTECFISPTSEQLTACGTNETYD  
SLTSTPSESGICRFTASPKPKDQGRVNMMAVPAKIAIGVTSYICLHTLEMTFC"

BASE COUNT  
1621 a 1658 c 1548 g 1283 t

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 20; DB 11; Length 6110;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgttgaccctgcaggaag 20  
DB 4560 CCTGTATGACTGACGAGAG 4579

RESULT 2  
LOCUS AF023450  
DEFINITION Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)  
ACCESSION AF023450  
VERSION AF023450.1 GI:3169767  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 6113)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a  
Down Syndrome Region and is Involved in the Development of the  
Nervous System  
Unpublished  
2 (bases 1 to 6113)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
Direct Submission  
Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research  
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los  
Angeles, CA 90048-1869, USA  
Location/Qualifiers  
1. 6413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21 (trisomy 21)"  
/map="21q22, between HMGI4 and MXI"  
/tissue\_type="brain"  
/dev\_stage="14 weeks, fetal"  
/clone="CHD2-52"  
/note="derived from alternately-spliced mRNA"  
1. 6413  
/gene="DSCAM"  
/note="member of immunoglobulin superfamily; involved in  
nervous system development"  
/product="Down syndrome cell adhesion molecule"  
/protein\_id="AAC17967.1"  
/db\_xref="GI:3169768"  
/translation="MMIALSLRPSFANVSEDLHSLYEYNASLOEVEFASTTGLV  
PCPAAGIPVTLRMVLATGEIYDVGIRVHPNGTLOIPPEPSSSTLIHNTYCTA  
ENSGKRSQDVHIRAVIREPTVREDOKTRMGVAVAKCIIPSSVEAFITVVEKDY  
DTSVSGSFLITSGALYIKDYNEDGLYRCITRRTTGETRQNSATLEPDSANAPSL  
IDFEDHRKMAGORVELPCKALGHEPDRLWKDNMPLSLGREFOKVTGLIENIRSD  
GSYCEVSNRGYTAIVIGRLVYKPLKATISPRKXSSGSOVSLSCVSTEDDELSTW  
YRNGELINPGKNRITGIGNHENLIMDHVKSDBGAYOCFYKDKLSADQVVOYVLE  
DTPKIIISAFSEKVVSPAEVSLMCNVKGPPLTITWTLDDPILGSGSHRISOMT  
SGNVSVLNTSSSOVDGQVYRGTANNSAGVLYOARINVRPASIRPKMTITAIAG  
RTYIHCYVIGYPIYSIKYKNSLPLPNROVAFENNGTLKLSQVKEVEDEEITCN  
VLYOPOLSTQSVHVTYKVPFTIOPEEPERSIQGVFICVAVSGDPLITTWQDGR  
IPGSLGVTIDNIDFTSLRIISLIMHNGNTICIAARNEAAVEHQSQLIVVPRFVQ  
RDDGIGYKAVILNCSAEGYPTIVYKFSKAGYQFOPALNGRIQVLSGLIKHVE  
EDGYGLCKVNSGADVSKSMILTVKIPAMITSYPTLATQSGKRESCAHGKRTI  
IRWEKEDRIINPEMAYLVSTKEGVEVITITOLIPVREDSGCEKATNSYGEDG  
ILQTOPEPDRPELEIKDKVAKRTITARMGPDGNSPTITGYIECKNSDSWDSQ  
RTKDYSPQNSKTIIDIPSSITSIRMTAKNRIGKSEPSNELITADBAADGPPOE  
VHLEPISSQIRVTWAPRKLQNGIIRGOYQIGREYSTGNGFONIIISVDGDSE  
VTLIDNLNKTQYGLVQACNRAGTGPSSOEITITLEDYSPSPENVAATSPESI  
SISWTSIKALNGILOGFVVIYMANIMDEGEIKNITTOPLSELDEGKXTNYSIO  
VLAFTRAGDGVSEIOFTRTKEDVGPAGVAKAASASAMVSMPLPKINGIIRKTY  
VCSHPYVUISEPASPSFSYRPNLSRNOVSANVAVNTSGRNSSEIITVERPLA  
KAPRILITSTYTTTMMKQIVLPCKAVGSPVAVMMKSNQSPSLVITDGRRIIS  
FSGSFIINTVAEDSGYSCIANNNWSDGDEILNLOVAPPDQPLVLSKTSSTLW  
LPDNGSSIRGYILIOYSEDSNEMGSPISPSERSYLENLCGTWKTLTQNGVGP  
RISEIIEAKLKEPOFSKOELFASINTRVNLIGNNDGCPITSFTLEVPFGTTV  
TAORTLSKSYLLYDQEAETELQMRVNSAGCEKAKRANTNYGSGTIPPLIKSV  
ONEGLTINELGKLMTSCILVGLVLLFVLLLYVARRRBR LKLRLRAKSLAEL  
MKNTRTSTLTLMKOOTLBMHIDIPRAOLILEERTWTITDRTVLLIDQDEAKO  
KSLIVTHVYOSVQATPLVDVSDAPGNPTTRNNAAGTANRLASQWTLNRPHTI  
SAHTLTIDWRLPTPRAAGSVKEDSYSPSODTBRARSNVSTESASTYEELARAE  
HAKMEOLRHAFTTECFISPTSEQLTACGTNETYD SLTSTPSESGICRFTASPK  
PKDQGRVNMMAVPAKIAIGVTSYICLHTLEMTFC"



```

/existence-not_experimental
11509..11621
/note="Xpound exon prediction, score = 74% (0%)"
/existence-not_experimental
11535..11621
/note="GRAIL, score = 58.000%, comment = good shadow"
/existence-not_experimental
12393..12480
/rpt_family="L2"
/existence-not_experimental
12492..12859
/rpt_family="L1M83"
/existence-not_experimental
13831..14128
/rpt_family="AlusX"
/existence-not_experimental
15597..15646
/rpt_family="L2"
/existence-not_experimental
15672..15819
/note="MZF, score = 53.3%"
/existence-not_experimental
15717..15762
/rpt_family="PABLA"
/existence-not_experimental
17207..17470
/rpt_family="AlusX"
/existence-not_experimental
17433..17534
/note="GRAIL, score = 87.000%, comment = excellent"
/existence-not_experimental
18789..19089
/rpt_family="AluY"
/existence-not_experimental
19490..20236
/rpt_family="L1PA15"
/existence-not_experimental
20244..26377
/rpt_family="L1PA2"
/existence-not_experimental
20451..22019
/note="GRAIL, score = 92.000%, comment = excellent"
/existence-not_experimental
21819..21970
/note="MZF, score = 65.5%"
/existence-not_experimental
22035..22267
/note="GRAIL, score = 59.000%, comment = good"
/existence-not_experimental
22537..22859
/note="GRAIL, score = 49.000%, comment = marginal"
/existence-not_experimental
23004..24272
/note="GRAIL, score = 87.000%, comment = excellent"
/existence-not_experimental
26405..27896
/rpt_family="L1PA15"
/existence-not_experimental
26551..26723
/note="GRAIL, score = 50.000%, comment = good"
/existence-not_experimental
26651..26773
/note="MZF, score = 52.1%"
/existence-not_experimental
27887..29472
/rpt_family="L1P"
/existence-not_experimental
28426..28580
/note="MZF, score = 63.6%"
/existence-not_experimental
29473..29835
/rpt_family="THE1B"
/existence-not_experimental

```

```

repeat_region 29836..29969
/rpt_family="L1P"
/existence-not_experimental
30121..30486
/rpt_family="THE1C"
/existence-not_experimental
30940..31299
/rpt_family="L1MC1"
/existence-not_experimental

Query Match 100.0%; Score 20; DB 11; Length 145861;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccggtatgacctgcaggaag 20
Db 56986 CCTGTATGACCTCGAGGAG 57005

RESULT 4
AF064865/c AF064865 159424 bp DNA PRI 02-JUN-1998
LOCUS Homo sapiens chromosome 21q22.3 PAC 58D10, complete sequence.
DEFINITION AF064865
ACCESSION AF064865.1 GI:3171160
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 159424)
AUTHORS Blechschmidt,K., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
TITLE Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
JOURNAL Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
Location/Qualifiers
source 1..159424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
complement(1224..1523)
/rpt_family="AluY"
/existence-not_experimental
1263..1386
/note="GC score = 8.60 (124bp)"
/note="Region: GC content"
/existence-not_experimental
1932..2047
/note="MZF, score = 92%"
/existence-not_experimental
3299..3348
/note="Xpound exon prediction, score = 72% (0%)"
/existence-not_experimental
3508..3614
/note="Genscan, score = -0.41%, comment = Terminal-exon
107 bp frame: 0 phase: 2"
/existence-not_experimental
3843..1124
/note="GRAIL, score = 53.000%, comment = good shadow"
/existence-not_experimental
3913..4018
/note="Genscan, score = 16.83%, comment = Initial-exon 106
bp frame: 1 phase: 1"
/existence-not_experimental
3913..4089
complement(3913..4089)
/note="Xpound exon prediction, score = 94% (0%)"
/existence-not_experimental
3913..4066
complement(3913..4066)
/note="MZF, score = 72%"
/existence-not_experimental
3913..4062
complement(3913..4062)

```





/evidence=not\_experimental  
26464. .26634

exon

Query Match 100.0%; Score 20; DB 11; Length 159424;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cctgtatgactgcaggaag 20  
|||||

Db 7672 CCTGTATGACTGCAGGAG 7653

RESULT 5  
AC013576/c  
LOCUS  
DEFINITION Homo sapiens chromosome 4 clone 364\_L\_4 map 4, LOW-PASS SEQUENCE  
AC013576 53718 bp DNA HTG 13-NOV-1999  
AC013576  
AC013576  
AC013576.1 GI:6403685  
HTG: HTGS\_PHASE0.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 53718)  
Biren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 4, clone 364\_L\_4  
Unpublished  
2 (bases 1 to 53718)  
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckery,R., Boguslavsky,L., Bouckhaalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardys,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lenoczky,J., Liu,C., Locke,K., MacDonald,P., Marquis,N.,  
McKwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.-J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 737: contig of 737 bp in length  
\* 738 1529: gap of unknown length  
\* 1530 2272: contig of 743 bp in length  
\* 2273 3041: gap of unknown length  
\* 3042 3795: gap of unknown length  
\* 3796 4567: contig of 772 bp in length  
\* 4568 5301: gap of unknown length  
\* 5301: contig of 734 bp in length

\* 5302 6044: gap of unknown length  
\* 6045 6809: gap of unknown length  
\* 6810 7559: contig of 750 bp in length  
\* 7560 8331: gap of unknown length  
\* 8332 9068: contig of 737 bp in length  
\* 9069 9894: gap of unknown length  
\* 9895 10600: contig of 706 bp in length  
\* 10601 11360: gap of unknown length  
\* 11361 12088: contig of 728 bp in length  
\* 12089 12904: gap of unknown length  
\* 12905 13665: contig of 761 bp in length  
\* 13666 14424: gap of unknown length  
\* 14425 15198: contig of 774 bp in length  
\* 15199 15951: gap of unknown length  
\* 15952 16705: contig of 754 bp in length  
\* 16706 17459: gap of unknown length  
\* 17460 18215: contig of 756 bp in length  
\* 18216 18996: gap of unknown length  
\* 18997 19777: contig of 781 bp in length  
\* 19778 20514: gap of unknown length  
\* 20515 21378: contig of 737 bp in length  
\* 21379 22154: gap of unknown length  
\* 22155 22915: contig of 775 bp in length  
\* 22916 23653: gap of unknown length  
\* 23654 24412: contig of 738 bp in length  
\* 24413 25203: gap of unknown length  
\* 25204 25979: contig of 791 bp in length  
\* 25980 26744: gap of unknown length  
\* 26745 27780: contig of 776 bp in length  
\* 27781 28447: gap of unknown length  
\* 28448 29203: contig of 765 bp in length  
\* 29204 29952: gap of unknown length  
\* 29953 30715: contig of 749 bp in length  
\* 30716 31477: gap of unknown length  
\* 31478 32218: contig of 763 bp in length  
\* 32219 32970: gap of unknown length  
\* 32970: contig of 762 bp in length  
\* 32971: gap of unknown length  
\* 32972: contig of 741 bp in length  
\* 32973: gap of unknown length  
\* 32974: contig of 752 bp in length  
\* 32975: gap of unknown length  
\* 32976: contig of 752 bp in length  
\* 32977: gap of unknown length  
\* 32978: contig of 752 bp in length  
\* 32979: gap of unknown length  
\* 32980: contig of 752 bp in length  
\* 32981: gap of unknown length  
\* 32982: contig of 752 bp in length  
\* 32983: gap of unknown length  
\* 32984: contig of 752 bp in length  
\* 32985: gap of unknown length  
\* 32986: contig of 752 bp in length  
\* 32987: gap of unknown length  
\* 32988: contig of 752 bp in length  
\* 32989: gap of unknown length  
\* 32990: contig of 752 bp in length  
\* 32991: gap of unknown length  
\* 32992: contig of 752 bp in length  
\* 32993: gap of unknown length  
\* 32994: contig of 752 bp in length  
\* 32995: gap of unknown length  
\* 32996: contig of 752 bp in length  
\* 32997: gap of unknown length  
\* 32998: contig of 752 bp in length  
\* 32999: gap of unknown length  
\* 33000: contig of 752 bp in length

```

* 32971 33745: contig of 775 bp in length
* 33746 34540: contig of 795 bp in length
* 34541 35368: contig of 828 bp in length
* 35369 36110: contig of 742 bp in length
* 36111 36883: contig of 773 bp in length
* 36884 37645: contig of 762 bp in length
* 37646 38401: contig of 756 bp in length
* 38402 39162: contig of 761 bp in length
* 39163 40020: contig of 858 bp in length
* 40021 40780: contig of 760 bp in length
* 40781 41521: contig of 741 bp in length
* 41522 42278: contig of 757 bp in length
* 42279 43046: contig of 768 bp in length
* 43047 43829: contig of 783 bp in length
* 43830 44587: contig of 758 bp in length
* 44588 45343: contig of 756 bp in length
* 45344 46097: contig of 754 bp in length
* 46098 46850: contig of 753 bp in length
* 46851 47613: contig of 763 bp in length
* 47614 48378: contig of 765 bp in length
* 48379 49134: contig of 756 bp in length
* 49135 49897: contig of 763 bp in length
* 49898 50661: contig of 764 bp in length
* 50662 51421: contig of 760 bp in length
* 51422 52189: contig of 768 bp in length
* 52190 52956: contig of 767 bp in length
* 52957 53718: contig of 762 bp in length.

```

```

FEATURES
  source
    1..53718
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="4"
      /map="4"
      /clone="364.L.4"
      /clone_lib="RPC1-11 Human Male BAC"

```

```

BASE COUNT 15526 a 11338 c 11023 g 15476 t 355 others
ORIGIN

```

```

Query Match 87.08: Score 17.4: DB 43: Length 53718:
Best Local Similarity 90.08: Pred. No. 60:
Matches 18: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

```

```

Qy 1 cctgtatgacctgcaggaag 20
    |||
Db 14737 CCTGTATGACCNACAGAG 14718

```

```

RESULT 6
LOCUS HS756G23/C
DEFINITION Human DNA sequence from clone 756G23 on chromosome 22q13.31-13.33,
complete sequence.
ACCESSION AL035681
VERSION AL035681.13 GI:4902689
KEYWORDS HTG: CPG Island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 89948)
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 27, 1999 this sequence version replaced gi:4775627.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM: EMBL; SW: SWISSPROT; TR: TREMBL
IMPORTANT: This sequence is not the entire insert of clone 756G23.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone dj979N1 is at 89849 in this sequence.
The true right end of clone 85F18 is at 80441 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An alt. pt is made to resolve all sequencing problems,
such as compressio. and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
feature key is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human Chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
756G23 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:pcrPAC2.

```

```

FEATURES
  source
    1..89948
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="22"
      /map="q13.31-13.33"
      /clone="RP4-756G23"
      /clone_lib="RPC1-4"

```

```

repeat_region 1..247
  /note="Alu repeat: matches 59. .296 of consensus"
repeat_region 350..571
  /note="MIR repeat: matches 12. .262 of consensus"
repeat_region 571..662
  /note="MIR repeat: matches 172. .262 of consensus"
repeat_region 663..948
  /note="Alu repeat: matches 1. .307 of consensus"
repeat_region 949..1069
  /note="MIR repeat: matches 47. .112 of consensus"
repeat_region 1084..1264
  /note="Alu repeat: matches 121. .309 of consensus"
repeat_region 1265..1557
  /note="Alu repeat: matches 1. .293 of consensus"
repeat_region 1558..1853
  /note="Alu repeat: matches 3. .300 of consensus"
repeat_region 1558..1873

```

```

/note="AluJb repeat: matches 118. .121 of consensus"
1889. .2013
repeat_region repeat: matches 62. .206 of consensus"
/note="MIR repeat: matches 62. .206 of consensus"
2172. .2466
repeat_region repeat: matches 1. .296 of consensus"
/note="AluSg1 repeat: matches 1. .296 of consensus"
2533. .2832
repeat_region repeat: matches 1. .297 of consensus"
/note="AluSx repeat: matches 1. .297 of consensus"
3566. .3743
repeat_region repeat: matches 143. .310 of consensus"
/note="AluSg repeat: matches 143. .310 of consensus"
3744. .4049
repeat_region repeat: matches 6. .311 of consensus"
/note="AluY repeat: matches 6. .311 of consensus"
4050. .4189
repeat_region repeat: matches 1. .143 of consensus"
/note="AluSg repeat: matches 1. .143 of consensus"
4628. .4930
repeat_region repeat: matches 1. .299 of consensus"
/note="AluJb repeat: matches 1. .299 of consensus"
4974. .5282
repeat_region repeat: matches 1. .308 of consensus"
/note="AluY repeat: matches 1. .308 of consensus"
5283. .5589
repeat_region repeat: matches 1. .313 of consensus"
/note="AluJb repeat: matches 1. .313 of consensus"
5731. .5815
repeat_region repeat: matches 20. .102 of consensus"
/note="MERSB repeat: matches 20. .102 of consensus"
5823. .6004
repeat_region repeat: matches 120. .301 of consensus"
/note="AluSg repeat: matches 120. .301 of consensus"
6005. .6039
repeat_region repeat: matches 11. .48 of consensus"
/note="MERSB repeat: matches 11. .48 of consensus"
6040. .6335
repeat_region repeat: matches 1. .302 of consensus"
/note="AluYb8 repeat: matches 1. .302 of consensus"
6363. .6663
repeat_region repeat: matches 1. .306 of consensus"
/note="AluYb8 repeat: matches 1. .306 of consensus"
6683. .6729
repeat_region repeat: matches 97. .145 of consensus"
/note="MIR repeat: matches 97. .145 of consensus"
6752. .7015
repeat_region repeat: matches 3. .299 of consensus"
/note="AluJb repeat: matches 3. .299 of consensus"
7580. .7755
repeat_region repeat: matches 7. .184 of consensus"
/note="MERSA repeat: matches 7. .184 of consensus"
7945. .8036
repeat_region repeat: matches 51. .139 of consensus"
/note="MIR repeat: matches 51. .139 of consensus"
8094. .8156
repeat_region repeat: matches 108. .178 of consensus"
/note="MERSB repeat: matches 108. .178 of consensus"
8157. .8413
repeat_region repeat: matches 1. .257 of consensus"
/note="AluSc repeat: matches 1. .257 of consensus"
8414. .8555
repeat_region repeat: matches 1. .108 of consensus"
/note="MERSB repeat: matches 1. .108 of consensus"
8556. .8677
repeat_region repeat: matches 107. .251 of consensus"
/note="MIR repeat: matches 107. .251 of consensus"
8735. .8878
repeat_region repeat: matches 55. .224 of consensus"
/note="MIR repeat: matches 55. .224 of consensus"
9384. .9485
repeat_region repeat: matches 86. .183 of consensus"
/note="MIR repeat: matches 86. .183 of consensus"
11657. .11940
repeat_region repeat: matches 1. .289 of consensus"
/note="AluSx repeat: matches 1. .289 of consensus"
12040. .12351
repeat_region repeat: matches 1. .308 of consensus"
/note="AluSg repeat: matches 1. .308 of consensus"
13156. .13462
repeat_region repeat: matches 1. .310 of consensus"
/note="AluSg repeat: matches 1. .310 of consensus"
13493. .13626
repeat_region repeat: matches 1. .135 of consensus"
/note="AluJb repeat: matches 1. .135 of consensus"
13832. .14106
repeat_region repeat: matches 1. .304 of consensus"
/note="AluY repeat: matches 1. .304 of consensus"
15009. .15035
repeat_region repeat: matches 2722. .2745 of consensus"
/note="L2 repeat: matches 2722. .2745 of consensus"
15036. .15328
repeat_region repeat: matches 1. .296 of consensus"
/note="AluY repeat: matches 1. .296 of consensus"
15329. .15628
repeat_region repeat: matches 1. .296 of consensus"
/note="AluSx repeat: matches 1. .296 of consensus"
15629. .15744
repeat_region repeat: matches 2588. .2722 of consensus"
/note="L2 repeat: matches 2588. .2722 of consensus"
15928. .16054
repeat_region repeat: matches 13. .133 of consensus"
/note="FLAMC repeat: matches 13. .133 of consensus"
16069. .16376
repeat_region repeat: matches 6. .309 of consensus"
/note="AluSx repeat: matches 6. .309 of consensus"
17747. .17885
repeat_region repeat: matches 48. .186 of consensus"
/note="MIR repeat: matches 48. .186 of consensus"

repeat_region 18933. .19033
/note="L2 repeat: matches 2575. .2678 of consensus"
repeat_region 19423. .19717
/note="AluSx repeat: matches 3. .297 of consensus"
repeat_region 21995. .22293
/note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 22402. .22687
/note="AluJb repeat: matches 1. .288 of consensus"
repeat_region 24842. .24969
/note="L2 repeat: matches 2529. .2680 of consensus"
repeat_region 25478. .25774
/note="AluY repeat: matches 2. .298 of consensus"
repeat_region 25642. .25938
/note="L2 r. eat: matches 2506. .2708 of consensus"
repeat_region 26437. .26601
/note="MIRIC repeat: matches 3. .173 of consensus"
repeat_region 26606. .26899
/note="AluSg repeat: matches 1. .294 of consensus"
repeat_region 26900. .27200
/note="MIRIC repeat: matches 173. .466 of consensus"
repeat_region 27798. .27974
/note="AluSc repeat: matches 138. .308 of consensus"
repeat_region 27975. .28278
/note="AluY repeat: matches 1. .294 of consensus"
repeat_region 28279. .28414
/note="AluSg repeat: matches 1. .138 of consensus"
repeat_region 29002. .29083
/note="MIR repeat: matches 65. .157 of consensus"
repeat_region 29080. .29296
/note="MIR repeat: matches 25. .262 of consensus"
repeat_region 29301. .29597
/note="AluSx repeat: matches 1. .297 of consensus"
repeat_region 29674. .29718
/note="MIR repeat: matches 107. .151 of consensus"
repeat_region 30167. .30254
/note="MIR repeat: matches 22. .114 of consensus"
repeat_region 30372. .30540
/note="L2 repeat: matches 2513. .2707 of consensus"
repeat_region 30943. .32264
/note="CPG island"
repeat_region 32358. .32573
/evidence-not-experimental
repeat_region 32700. .32845
/note="L2 repeat: matches 2370. .2624 of consensus"
repeat_region 33210. .33276
/note="MIR repeat: matches 1. .152 of consensus"
repeat_region 33500. .33794
/note="MIR repeat: matches 107. .171 of consensus"
repeat_region 34840. .35023
/note="AluJb repeat: matches 23. .305 of consensus"
repeat_region 35161. .35287
/note="L1M1 repeat: matches 5687. .5884 of consensus"
repeat_region 35288. .35562
/note="AluJb/FLAM repeat: matches 1. .131 of consensus"
repeat_region 35563. .35664
/note="AluY repeat: matches 1. .296 of consensus"
repeat_region 35563. .35664
/note="51 copies 2 mer ta 86% conserved"

Query Match 87.0% Score 17.4: DB 11: Length 89948;
Best Local Similarity 94.7% Pred. No. 63:
Matches 18: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 cgtatgacctgcagaag 20
DB 1038 ctttatgacctgcagaag 1020

RESULT 7
HS74714 106319 bp DNA PRI 23-NOV-1999
LOCUS HS74714
DEFINITION Homo sapiens DNA sequence from PAC 74714 on chromosome 1q23-24.
Contains a Brachyury (T box protein)-LIKE gene, ESTs, STSs and a CA
repeat polymorphism.

```

ACCESSION AL009051  
VERSION AL009051.1 GI:2995195  
KEYWORDS Brachyury; CA repeat; T box.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106319)  
AUTHORS Wray, P.  
TITLE Direct Submission  
JOURNAL Submitted (11-MAR-1998) sanger.ac.uk/HGP/Chr1/ Sanger Centre,  
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humuery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Mar 28, 1998 this sequence version replaced gi:2644985.  
IMPORTANT: This sequence is the entire insert of clone 747L4.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variations annotated may not be found in the sequence submission  
corresponding to the overlapping clone as we submit sequences with  
only a small overlap as described above.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 1, constructed by the Sanger Centre chromosome 1  
mapping group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1/  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsur.'  
feature key.  
The true right end of clone 747L4 is at 1 in this sequence. The  
true right end of this clone is at 106319.  
747L4 is from the library RPC14 constructed at the Roswell Park  
Cancer Institute by the group of Pietri de Jong.  
For further details see http://bacpac.med.buffalo.edu/  
Location/Qualifiers  
1..106319  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q23-q24"  
/clone="RP4-747L4"  
/clone\_1lb="RPC1-4"  
1..83  
/note="L1MA5A repeat: matches 960. .1042 of consensus"  
repeat\_region  
1067..1370  
/note="Alus repeat: matches 13. .302 of consensus"  
repeat\_region  
3124..3228  
/note="MIR repeat: matches 66. .168 of consensus"  
repeat\_region  
3236..3687  
/note="L1MA7 repeat: matches 457. .920 of consensus"  
repeat\_region  
3705..4005  
/note="Alus repeat: matches 3. .302 of consensus"  
repeat\_region  
5016..5049  
/note="17 copies of 2 mer 94 & conserved"  
repeat\_region  
5280..5367  
/note="LRR8 repeat: matches 474. .390 of consensus"  
repeat\_region  
5654..5730  
/note="LRR8 repeat: matches 84. .1 of consensus"  
repeat\_region  
5900..6011  
/note="2 copies of 56 mer 100 & conserved"  
repeat\_region  
8036..8199  
/note="MERS5 repeat: matches 154. .2 of consensus"  
repeat\_region  
8399..8524  
/note="MIR2 repeat: matches 142. .14 of consensus"  
repeat\_region  
8402..8900  
/note="match: STS G35325"  
misc\_feature  
complement(8402..9600)  
/note="match: ESTs N3939 W03699"  
prim\_transcript  
8892..9193  
/note="Alu repeat: matches 302. .3 of consensus"  
repeat\_region  
complement(9859..41729)  
gene

CDS  
/gene="dJ747L4.1"  
complement(join(<9859..10118,13949..14084,17630..17818,  
22343..22404,25041..25102,29548..29682,31402..31666,  
41533..41729))  
/gene="dJ747L4.1"  
/note="match: proteins O15178 P79778 P24781 Q25113 P79777  
P20293 O02737 Q42100 O07998 P80492 Q17134 P55158 P55965;  
match: cDNAs U67087 X51683 U25176 AB001871 AJ001699 D89442  
S57147 M77243 X91903 D50332 S74163 D16441; match: genomic  
DNA AJ001528"  
/codon\_start=1  
/evidence=not-experimental  
/product="dJ747L4.1 (Brachyury (T box protein))-LIKE  
protein"  
/protein\_id="CAA15624.1"  
/db\_xref="GI:3900891"  
/db\_xref="SPTREMBL:O60806"  
/translation="MSLGTRKPSGTYSHLNYVESELOAGREKDPTEKQDITILE  
DAPLWQREKEVINEMIYTRNGRRMPVLKISTVGLDPAKMSLIDFPTSHRKTY  
NDEWPAKREVSASHSCVYIPDPSNFGAHMKAPISFSKYALKLNGGQIMLSL  
HKYEPOHIVRGSAHRMVTNCSPEPTQFIATAYONEIATLKIYMPFAKFLDAK  
ERNHLDPDEALSESOHVTYSLGGMIFSNPDGCTAGNSNYQAAPLPAPPHHG  
CEHYSGLRGHROAPPSAYMHRNHSVNLIESNNLOVPSGPDSTSLSTPSASI  
LSYPRHNCIPNCPSPRYCMTIISNGACRPSGCEVHASPGLLGNPAVTSPPSV  
LSIQAPTSAGVEVLDPSSLTIAVSTWIAVASHPRAGMGPEA"  
10517..10722  
/note="MIR repeat: matches 38. .260 of consensus"  
10852..11019  
/note="MERS5 repeat: matches 157. .11 of consensus"  
11455..11593  
/note="MIR repeat: matches 63. .202 of consensus"  
11784..11995  
/note="MERS3 repeat: matches 302. .82 of consensus"  
12103..12303  
/note="MERS3 repeat: matches 228. .1 of consensus"  
12224..12433  
/note="MERS5A repeat: matches 30. .163 of consensus"  
12516..12806  
/note="Alus repeat: matches 297. .1 of consensus"  
12873..13117  
/note="MIR repeat: matches 261. .13 of consensus"  
13435..13735  
/note="Alus repeat: matches 301. .1 of consensus"  
14546..14679  
/note="FLAM\_C repeat: matches 1. .133 of consensus"  
14705..15083  
/note="L1MA8 repeat: matches 639. .1019 of consensus"  
15027..15040  
/gene="dJ747L4.1"  
15238..15616  
/note="MSTD repeat: matches 1. .394 of consensus"  
16054..16110  
/note="MERSA repeat: matches 91. .35 of consensus"  
16180..16480  
/note="Alus repeat: matches 303. .1 of consensus"  
16490..16565  
/note="MIR2 repeat: matches 60. .135 of consensus"  
16817..16945  
/note="MIR2 repeat: matches 4. .144 of consensus"  
18036..18170  
/note="LRR8 repeat: matches 1. .146 of consensus"  
18360..18686  
/note="LRR8 repeat: matches 310. .625 of consensus"  
19428..19773  
/note="L1PA5 repeat: matches 542. .890 of consensus"  
21266..21557  
/note="Alus repeat: matches 1. .292 of consensus"  
23054..23348  
/note="Alu repeat: matches 2. .299 of consensus"  
23609..23894  
/note="Alus repeat: matches 3. .302 of consensus"  
23876..24232  
/gene="dJ747L4.1"

```

repeat_region 24617..24911
/Note="AluY repeat: matches 295..1 of consensus"
repeat_region 25217..25267
/Note="AluSx/g repeat: matches 3..53 of consensus;
incomplete repeat"
repeat_region 25460..25555
/Note="MIR2 repeat: matches 137..37 of consensus"
repeat_region 25767..25911
/Note="MIR2 repeat: matches 2..146 of consensus"
repeat_region 25950..26099
/Note="AluSg repeat: matches 288..134 of consensus;
incomplete repeat"
repeat_region 26165..26522
/Note="THEIC repeat: matches 2..371 of consensus"
repeat_region 27520..27815
/Note="AluSg repeat: matches 296..1 of consensus"
repeat_region 28025..28325
/Note="AluSx repeat: matches 1..301 of consensus"
repeat_region 28326..28492
/Note="AluSg repeat: matches 134..300 of consensus;
incomplete repeat"
repeat_region 28497..28589
/Note="LIME2 repeat: matches 478..568 of consensus"
repeat_region 29500..29541
/Note="21 copies of 2 mer 100 & conserved"
repeat_region 29943..30242
/Note="AluY repeat: matches 301..2 of consensus"
repeat_region 32073..32270
/Note="AluJo repeat: matches 85..295 of consensus;
incomplete repeat"
repeat_region 32597..32662
/Note="Limb6 repeat: matches 812..747 of consensus"
repeat_region 32688..32914
/Note="AluSx repeat: matches 224..1 of consensus;
incomplete repeat"
repeat_region 33089..33190
/Note="MIR repeat: matches 260..162 of consensus"
repeat_region 33838..34067
/Note="Aluub repeat: matches 73..302 of consensus;
incomplete repeat"
repeat_region 34284..34400
/Note="FLAM_C repeat: matches 131..6 of consensus"
repeat_region 34443..34518
/Note="MER2 repeat: matches 345..268 of consensus"
repeat_region 34546..34752
/Note="MER2 repeat: matches 210..1 of consensus"
repeat_region 34840..34962
/Note="AluSg repeat: matches 134..300 of consensus;
incomplete repeat"
repeat_region 34845..34888
/Note="MIR repeat: matches 152..109 of consensus"

Query Match 87.0%: Score 17.4; DB 11; Length 106319;
Best Local Similarity 94.7%: Pred. No. 65;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cctgtatgacctgcagaa 19
|||||
DB 90330 CCTGATGAC TCGAGGA 90312

RESULT 8
AC009389 116967 bp DNA HTG 20-SEP-1999
LOCUS AC009389
DEFINITION Drosophila melanogaster chromosome 3 clone BACR01F12 (D1005)
RPCI-98 01.F.12 map 98C-98C strain Y; cn bw sp, *** SEQUENCING IN
PROCESSES ***, 93 unordered pieces.
AC009389 AC009389.5 GI:5912649
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

REFERENCE
AUTHORS
Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 116967)
Celisner,S.E., Abgaryan,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 116967)
Celisner,S.E., Abgaryan,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (20-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 20, 1999 this sequence version replaced g1:5832523.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bd@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a "working draft" sequence. It currently
* consists of 93 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 747: contig of 747 bp in length
748 827: gap of unknown length
828 1318: contig of 491 bp in length
1319 1398: gap of unknown length
1399 1986: contig of 588 bp in length
1987 2066: gap of unknown length
2067 2641: contig of 575 bp in length
2642 2722: gap of unknown length
2722 3329: contig of 558 bp in length
3329 3359: gap of unknown length
3359 3280: gap of unknown length
3280 3360: gap of unknown length
3360 3817: contig of 438 bp in length
3818 3897: gap of unknown length
3898 4897: contig of 1000 bp in length
4898 4977: gap of unknown length
4978 5629: contig of 652 bp in length
5630 5709: gap of unknown length
5710 6344: contig of 635 bp in length
6345 6425: gap of unknown length
6426 6934: contig of 510 bp in length
6935 7015: gap of unknown length
7016 7911: contig of 897 bp in length
7912 7992: gap of unknown length
7993 8630: contig of 639 bp in length
8631 8710: gap of unknown length
8711 9259: contig of 549 bp in length
9260 9339: gap of unknown length
9340 10004: contig of 665 bp in length
10005 10084: gap of unknown length
10085 10869: contig of 785 bp in length
10870 10949: gap of unknown length
10950 11505: contig of 536 bp in length
11506 11585: gap of unknown length
11586 12603: contig of 1018 bp in length

```

```

* 12604 12683: gap of unknown length
* 12684 13677: contig of 994 bp in length
* 13678 13757: gap of unknown length
* 13758 14648: contig of 891 bp in length
* 14649 14728: gap of unknown length
* 14729 15864: contig of 1136 bp in length
* 15865 15944: gap of unknown length
* 15945 16668: contig of 924 bp in length
* 16669 16948: gap of unknown length
* 16949 17912: contig of 964 bp in length
* 17913 17992: gap of unknown length
* 17993 18748: contig of 756 bp in length
* 18749 18828: gap of unknown length
* 18829 19541: contig of 713 bp in length
* 19542 19622: gap of unknown length
* 19623 21172: contig of 1551 bp in length
* 21173 21252: gap of unknown length
* 21253 22034: contig of 782 bp in length
* 22035 22115: gap of unknown length
* 22116 22996: contig of 882 bp in length
* 22997 23076: gap of unknown length
* 23077 24183: contig of 1107 bp in length
* 24184 24263: gap of unknown length
* 24264 24846: contig of 583 bp in length
* 24847 24926: gap of unknown length
* 24927 26566: contig of 1640 bp in length
* 26567 27522: gap of unknown length
* 27523 27602: contig of 876 bp in length
* 27603 28577: gap of unknown length
* 28578 28657: contig of 975 bp in length
* 28658 29778: gap of unknown length
* 29779 29858: contig of 1121 bp in length
* 29859 30517: gap of 659 bp in length
* 30518 30957: gap of unknown length
* 30958 31572: contig of 975 bp in length
* 31573 31652: gap of unknown length
* 31653 32385: contig of 733 bp in length
* 32386 32465: gap of unknown length
* 32466 33129: contig of 664 bp in length
* 33130 33209: gap of unknown length
* 33210 34414: contig of 1205 bp in length
* 34415 34494: gap of unknown length
* 34495 35703: contig of 1209 bp in length
* 35704 35783: gap of unknown length
* 35784 36586: contig of 873 bp in length
* 36587 36736: gap of unknown length
* 36737 37871: contig of 1135 bp in length
* 37872 37951: gap of unknown length
* 37952 39259: contig of 1308 bp in length
* 39260 39339: gap of unknown length
* 39340 39956: contig of 617 bp in length
* 39957 40036: gap of unknown length
* 40037 41349: contig of 1313 bp in length
* 41350 41429: gap of unknown length
* 41430 42775: contig of 1346 bp in length
* 42776 42855: gap of unknown length
* 42856 44226: contig of 1371 bp in length
* 44227 44306: gap of unknown length
* 44307 44942: contig of 636 bp in length
* 44943 45022: gap of unknown length
* 45023 46435: contig of 1413 bp in length
* 46436 46515: gap of unknown length
* 46516 47405: contig of 890 bp in length
* 47406 47485: gap of unknown length
* 47486 49222: contig of 1737 bp in length
* 49223 49302: gap of unknown length
* 49303 50123: contig of 821 bp in length
* 50124 50203: gap of unknown length
* 50204 51739: contig of 1536 bp in length
* 51740 51819: gap of unknown length
* 51820 53455: contig of 1636 bp in length
* 53456 53535: gap of unknown length

```

```

* 53536 55569: contig of 2034 bp in length
* 55570 55649: gap of unknown length
* 55650 57240: contig of 1591 bp in length
* 57241 57320: gap of unknown length
* 57321 58691: contig of 1371 bp in length
* 58692 58771: gap of unknown length
* 58772 60696: contig of 1925 bp in length
* 60697 60776: gap of unknown length
* 60777 62740: contig of 1964 bp in length
* 62741 62820: gap of unknown length
* 62821 64463: contig of 1643 bp in length
* 64464 64543: gap of unknown length
* 64544 67207: contig of 2664 bp in length
* 67208 67287: gap of unknown length
* 67288 69092: contig of 1805 bp in length
* 69093 69172: gap of unknown length
* 69173 72712: contig of 3540 bp in length
* 72713 72792: gap of unknown length
* 72793 74880: contig of 2088 bp in length
* 74881 74960: gap of unknown length
* 74961 77394: contig of 2434 bp in length
* 77395 77474: gap of unknown length
* 77475 80512: contig of 3038 bp in length
* 80513 80592: gap of unknown length
* 80593 83793: contig of 3121 bp in length
* 83794 88050: contig of 4257 bp in length
* 88051 88130: gap of unknown length
* 88131 90875: contig of 2745 bp in length
* 90876 90955: gap of unknown length
* 90956 95358: contig of 4403 bp in length
* 95359 95438: gap of unknown length
* 95439 101482: contig of 6044 bp in length
* 101483 101562: gap of unknown length
* 101563 102222: contig of 660 bp in length
* 102223 102302: gap of unknown length
* 102303 102995: contig of 693 bp in length
* 102996 103075: gap of unknown length
* 103076 103600: contig of 525 bp in length
* 103601 103680: gap of unknown length
* 103681 104345: contig of 665 bp in length
* 104346 104425: gap of unknown length
* 104426 104975: contig of 550 bp in length

Query Match      87.0%: Score 17.4; DB 41; Length 116967;
Best Local Similarity 94.7%: Pred. No. 65;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 cctgtatgacctgcagaa 19
Db 19056 CCTGTATGCCCTGCAGAA 19074

RESULT 9
AC011140 139134 bp DNA HTG 08-DEC-1999
LOCUS Homo sapiens clone RP11-116, *** SEQUENCING IN PROGRESS ***, 23
DEFINITION unordered pieces.
ACCESSION AC011140
VERSION AC011140.2 GI:6539376
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 139134)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-116
REFERENCE 2 (bases 1 to 139134)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslajsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

```

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Funk, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Lehoczy, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,  
 McEwan, P., McGuire, A., McKernan, K., McLaughlin, J., Meldrum, J.,  
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Tirell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Whitehead Institute/MIT Center for Genome Research  
 Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 8, 1999 this sequence version replaced gi:6006163.

All repeats were identified using RepeatMasker:

Smt, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- genome center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project information

Center project name: 1\_1-6

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1297: contig of 1297 bp in length  
 \* gap of unknown length  
 \* 1298 2645: contig of 1348 bp in length  
 \* gap of unknown length  
 \* 2646 4193: contig of 1548 bp in length  
 \* gap of unknown length  
 \* 4194 5882: contig of 1689 bp in length  
 \* gap of unknown length  
 \* 5883 7953: contig of 2071 bp in length  
 \* gap of unknown length  
 \* 7954 9466: contig of 1513 bp in length  
 \* gap of unknown length  
 \* 9467 12511: contig of 3045 bp in length  
 \* gap of unknown length  
 \* 12512 15375: contig of 2864 bp in length  
 \* gap of unknown length  
 \* 15376 18067: contig of 2692 bp in length  
 \* gap of unknown length  
 \* 18068 22522: contig of 4455 bp in length  
 \* gap of unknown length  
 \* 22523 28254: contig of 5732 bp in length  
 \* gap of unknown length  
 \* 28255 32322: contig of 4068 bp in length  
 \* gap of unknown length  
 \* 32323 38431: contig of 6109 bp in length  
 \* gap of unknown length  
 \* 38432 43919: contig of 5488 bp in length  
 \* gap of unknown length  
 \* 43920 51206: contig of 7287 bp in length  
 \* gap of unknown length  
 \* 51207 58769: contig of 7563 bp in length  
 \* gap of unknown length  
 \* 58770 69046: contig of 10277 bp in length  
 \* gap of unknown length  
 \* 69047 80029: contig of 10983 bp in length  
 \* gap of unknown length  
 \* 80030 88523: contig of 8494 bp in length  
 \* gap of unknown length

## FEATURES

source

\* 88524 98308: contig of 9785 bp in length  
 \* gap of unknown length  
 \* 98309 112715: contig of 14407 bp in length  
 \* gap of unknown length  
 \* 112716 124886: contig of 12171 bp in length  
 \* gap of unknown length  
 \* 124887 139134: contig of 14248 bp in length.  
 Location/Qualifiers  
 1. 139134 "Homo sapiens"  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="RP11-116"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 32063 c 29598 g 38453 t 23 others

## BASE COUNT

38997 a 32063 c 29598 g 38453 t 23 others

## ORIGIN

Query Match 85.0%; Score 17; DB 45; Length 139134;  
 Best Local Similarity 100.0%; Pred. No. 1; e=0; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctgtatgacctgcag 17

Db 64302 CCTGTATGACCTGCAG 64318

## RESULT 10

HSMYOH

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

CDS

1. 3374

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/dev\_stage="fetal"

/tissue\_type="cardiac muscle"

/clone\_id="lambda gt10"

/clone\_lib="psmHCP"

&lt;1. 3337

/product="myosin heavy chain (1078 AA)"

/protein\_id="CA68757.1"

/db\_xref="GI:34864"

/db\_xref="SWISS-PROT:P13535"

/translation="OKTDELAKSEARKKELEKMYLLIKKNDLOQOVSEADSLAD  
 AEERPOIKKKIOLEAKIKETVERAEKEEINELVAKRKLEDECSLEKIDDE  
 LTLAVBOKKATENKVKNLKEEMAGLDETIAKLSKEKALQETHOOTIDMENDKO  
 VILTKATKLEQVDLEGLDEOKKLRMDLERAKRKLEEDLKLQBSTYDNRARAKA  
 QDEKLEKKEFEISNLSKIDEDFOKVELOKRIKELQARIEEGEIEARASRA  
 EKORSDIKLEIEISERLEAGATSAOVELINKREAFOKLRDLEATLQHEAWA  
 ALRKHALSMALGEQIDNLOVKOKLEKSEKLEKMEFDKSSNAEISAKSGHLEK  
 CSLEQVGLKTEKEEQRLINDLTARARLQTAGESRLIDEDALVQLSKSKO  
 ASTQOIEHKLHLEETKAKNALAHALSSRDCLLRQYIEEBEGKAELOARLSKA  
 NSSEAVMRKRIEIDARTLELEAKKILAQRLQDEEHVEVNNKCSALPCTKORLO  
 NEVEDIMLDEVERSNACALDKKORNFQVLSKKOKVEETOAELEASOKESRLSTE











```

REFERENCE 1 (bases 1 to 3513)
AUTHORS Smith,J.B. and Herschman,H.R.
TITLE The glucocorticoid attenuated response genes GARG-16, GARG-39, and
GARG-49/ING2 encode inducible proteins containing multiple
tetraatricopeptide repeat domains
JOURNAL Arch. Biochem. Biophys. 330 (2), 290-300 (1996)
MEDLINE 96239145
REFERENCE 2 (bases 1 to 3513)
AUTHORS Smith,J.B. and Herschman,H.R.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1995) Jeffrey B. Smith, Pediatrics, UCLA School
of Medicine, 10833 Le Conte Avenue, Los Angeles, CA 90095-1752, USA
FEATURES
Source Location/Qualifiers
1..3513
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="Swiss 3T3"
52..1464
/gene="GARG-39"
52..1464
/gene="GARG-39"
/standard_name="glucocorticoid-attenuated response gene
39"
/note="IFIT-2: contains ten tetraatricopeptide repeat (TPR)
domains: homolog of human ISG-54K"
/codon_start=1
/product="GARG-39"
/protein_id="AAC52626.1"
/db_xref="GI:1401062"
/translation="MSTTSKESLESNLRQKCHPTWNLIAEDESLEDFEDRFVFNKDEF
ONSEKATMNCINLAYKHCRCIGENALQCLGEAEGLIOOHDPDVEIESTLVATGNVAV
VYHMGFSKQAVLYDKVKQCKKSSPYRIENPALDCEEGMARLCKTKNNEVKYC
FOKALEKPKNPETSGAIANYRIIDPMPARJYCDSDSEAOIOLSPNTYKYVLTAIK
LDVHKNQMALVEALRKKDSALDTLLRARFYKYKYDTPRALIOLRKALEKTPNNA
VYHYVGGCCYRSKRVHMLNRREWVPSGDRKKLEELIQAVNHLKAEIKETKESCS
FLADLYIAKKYDEADYVFKELSKDLPGRQQLHLRYGNFOFKMRQOKAIYHYM
EGVKIKKTIIPQKREKRLQRLRLRHLHEDSDSEALHILFLQENGSGQADQDSER
CVDSANOVPSASLDEGAAY"
203..304
/gene="GARG-39"
/note="encodes TPR 1 domain"
312..433
/gene="GARG-39"
/note="encodes TPR 2 domain"
521..565
/gene="GARG-39"
/note="encodes TPR 3b domain"
575..676
/gene="GARG-39"
/note="encodes TPR 4 domain"
677..775
/gene="GARG-39"
/note="encodes TPR 5 domain"
776..877
/gene="GARG-39"
/note="encodes TPR 6 domain"
878..910
/gene="GARG-39"
/note="encodes TPR 7a domain"
995..1039
/gene="GARG-39"
/note="encodes TPR 7b domain"
1040..1141
/gene="GARG-39"
/note="encodes TPR 8 domain"
1151..1255
/gene="GARG-39"
/note="encodes TPR 9 domain"
1328..1426
/gene="GARG-39"
/note="encodes TPR 10 domain"
misc_count 1130 a 641 c 792 g 950 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 12; Length 3513;
Best Local Similarity 90.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 ccagttctcaagagcaag 20
Db 375 CCACTTCTCAAGGCTCAGG 394
RESULT 15
AC008293
LOCUS
DEFINITION
AC008293 55741 bp DNA HTG 02-AUG-1999
Drosophila melanogaster chromosome 3 clone DS01181 (0512) map
89D-89D strain y: cn bw sp, *** SEQUENCING IN PROGRESS ***, 40
unordered pieces.
AC008293 GI:5670673
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 55741)
Celisner,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomoclan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paaleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequiera,A., Sethi,H., Smlr,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 55741)
Celisner,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomoclan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paaleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequiera,A., Sethi,H., Smlr,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
13-29.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1095: contig of 1095 bp in length
* 1096 1175: gap of unknown length
* 1176 2015: contig of 840 bp in length
* 2016 2095: gap of unknown length
* 2096 2737: contig of 642 bp in length
* 2738 2817: gap of unknown length
* 2818 4076: contig of 1259 bp in length
* 4077 4156: gap of unknown length
* 4157 4991: contig of 835 bp in length

```

## COMMENT

```
* 4992 5071: gap of unknown length
* 5072 5668: contig of 597 bp in length
* 5669 5748: gap of unknown length
* 5749 6979: contig of 1231 bp in length
* 6980 7059: gap of unknown length
* 7060 7950: contig of 891 bp in length
* 7951 8030: gap of unknown length
* 8031 8745: contig of 715 bp in length
* 8746 8825: gap of unknown length
* 8826 9779: contig of 954 bp in length
* 9780 9860: gap of unknown length
* 9860 10832: contig of 973 bp in length
* 10833 10912: gap of unknown length
* 10913 12139: contig of 1227 bp in length
* 12140 12219: gap of unknown length
* 12220 13632: contig of 1413 bp in length
* 13633 13712: gap of unknown length
* 13713 15303: contig of 1591 bp in length
* 15304 15383: gap of unknown length
* 15384 16385: contig of 1002 bp in length
* 16386 16465: gap of unknown length
* 16466 17393: contig of 928 bp in length
* 17394 17473: gap of unknown length
* 17474 18788: contig of 1315 bp in length
* 18789 18868: gap of unknown length
* 18869 20618: contig of 1750 bp in length
* 20619 20698: gap of unknown length
* 20699 21896: contig of 1198 bp in length
* 21897 21976: gap of unknown length
* 21977 23559: contig of 1583 bp in length
* 23560 23639: gap of unknown length
* 23640 25202: contig of 1563 bp in length
* 25203 25282: gap of unknown length
* 25283 26948: contig of 1666 bp in length
* 26949 27028: gap of unknown length
* 27029 28616: contig of 1588 bp in length
* 28617 28696: gap of unknown length
* 28697 30859: contig of 2163 bp in length
* 30860 30939: gap of unknown length
* 30940 32463: contig of 1524 bp in length
* 32464 32543: gap of unknown length
* 32544 34287: contig of 1744 bp in length
* 34288 34367: gap of unknown length
* 34368 37264: contig of 2897 bp in length
* 37265 37344: gap of unknown length
* 37345 40346: contig of 3002 bp in length
* 40347 40426: gap of unknown length
* 40427 45018: contig of 4592 bp in length
* 45019 45098: gap of unknown length
* 45099 48778: contig of 3680 bp in length
* 48779 48858: gap of unknown length
* 48859 49532: contig of 694 bp in length
* 49533 49632: gap of unknown length
* 49634 50244: contig of 612 bp in length
* 50245 50324: gap of unknown length
* 50325 51056: contig of 732 bp in length
* 51057 51135: gap of unknown length
* 51137 51792: contig of 656 bp in length
* 51793 51872: gap of unknown length
* 51873 52214: contig of 342 bp in length
* 52215 52294: gap of unknown length
* 52295 53035: contig of 651 bp in length
* 52956 53673: gap of unknown length
* 53036 53674: contig of 638 bp in length
* 53675 53753: gap of unknown length
* 53754 54282: contig of 529 bp in length
* 54283 54362: gap of unknown length
* 54363 55053: contig of 691 bp in length
* 55054 55133: gap of unknown length
* 55134 55741: contig of 608 bp in length.
```

## FEATURES

SOURCE

```
Location/Qualifiers
1..55741
/organism="Drosophila melanogaster"
```

```
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="89D-89D"
```

```
/clone="DS01181 (D512)"
/clone_1lb="PI library, partial Sau3A in pMS582vec14Ad10"
```

```
BASE COUNT      14707 a 11367 c 11512 g 10032 t 3123 others
```

ORIGIN

## Query Match

84.0%: Score 16.8; DB 33: Length 55741;

Best Local Similarity 90.0%: Pred. No. 85;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccagttctcaagagcagc 20

||| ||||| ||||| |||

Db 20591 CCATTCTCAAGAGCCCG 20610

Search completed: April 25, 2000, 20:38:36

Job time: 21124 sec

27

LOCUS	HSDD468N4	237415 bp	DNA	HTG	23-NOV-1999
DEFINITION	Homo sapiens chromosome 20 clone RP3-468N4, *** SEQUENCING IN PROGRESS ***, in unordered pieces.				
ACCESSION	AL121752				
VERSION	AL121752.2	GI:6136980			
KEYWORDS	HTG: HTGS_PHASE1.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 237415)				
JOURNAL	Steward,C.				
COMMENT	Direct Submission Submitted (26-OCT-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Oct 27, 1999 this sequence version replaced g16014391. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: dj468N4 Contig_ID: 00733 acc-AL121752 Length: 1449 bp Unfinished: dj468N4 Contig_ID: 01553 acc-AL121752 Length: 36096 bp Unfinished: dj468N4 Contig_ID: 01594 acc-AL121752 Length: 35200 bp Unfinished: dj468N4 Contig_ID: 01533 acc-AL121752 Length: 107601 bp Unfinished: dj468N4 Contig_ID: 01678 acc-AL121752 Length: 39960 bp Unfinished: dj468N4 Contig_ID: 01724 acc-AL121752 Length: 11109 bp. * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.				
FEATURES	Location/Qualifiers				
source	1..237415 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20" /clone="RP3-468N4" /clone_1fb="RPC1-3"				
BASE COUNT	65944 a 53184 c 52924 g 61362 t 4001 others				
ORIGIN					
Query Match	85.0%; Score 17; DB 32; Length 237415:				
Best Local Similarity	100.0%; Pred .. 07:				
Matches	17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
0Y	4 gttctcaagaagcagcag 20       				
Db 102945	GTTCTCMAAGAGCAGC 102929				
RESULT 11					
LOCUS	S77710S2	1742 bp	DNA	ROD	27-SEP-1995
DEFINITION	Ifi55-interferon-alpha regulated gene [mice, Genomic, 1742 nt, segment 2 of 2].				
ACCESSION	S77713				
VERSION	S77713.1	GI:998614			
KEYWORDS	2 of 2				
SEGMENT	Mus sp.				
SOURCE	Mus sp.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukaryota; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1742)				
AUTHORS	Bluyssen,H.A., Vlietstra,R.J., Faber,P.W., Smit,E.M., Hagemeljer,A. and Trapman,J.				
TITLE	Structure, chromosome localization, and regulation of expression of the interferon-regulated mouse Ifi554/Ifi56 gene family				
JOURNAL	Genomics 24 (1), 137-148 (1994)				

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:17 ; Search time 7810.03 Seconds  
(without alignments)  
-7.776 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20  
Sequence: 1 ccaagttccaaagagcagcag 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_dal: \*  
2: gb\_baz: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_p11: \*  
8: gb\_p12: \*  
9: gb\_p1: \*  
10: gb\_p12: \*  
11: gb\_p13: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_v1: \*  
17: em\_fun: \*  
18: em\_huml: \*  
19: em\_hum2: \*  
20: em\_in: \*  
21: em\_om: \*  
22: em\_or: \*  
23: em\_ov: \*  
24: em\_pat: \*  
25: em\_ph: \*  
26: em\_p1: \*  
27: em\_ro: \*  
28: em\_sts: \*  
29: em\_sy: \*  
30: em\_un: \*  
31: em\_v1: \*  
32: gb\_htg1: \*  
33: gb\_htg2: \*  
34: gb\_in1: \*  
35: gb\_in2: \*  
36: em\_ba1: \*  
37: em\_ba2: \*  
38: em\_hum3: \*  
39: em\_hum4: \*  
40: gb\_p14: \*  
41: gb\_htg3: \*  
42: gb\_htg4: \*  
43: gb\_htg5: \*  
44: gb\_htg6: \*

45: gb\_htg7: \*  
46: em\_htg1: \*  
47: em\_htg2: \*  
48: em\_htg3: \*  
49: em\_hum5: \*  
50: gb\_p13: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	6110	11 AF023449	AF023449 Homo sapi
2	20	100.0	6413	11 AF023450	AF023450 Homo sapi
3	20	100.0	145861	11 AF064862	AF064862 Homo sapi
4	20	100.0	159424	11 AF064865	AF064865 Homo sapi
5	17.4	87.0	3654	4 CHKPI05A	M86930 Chicken p10
6	17.4	87.0	3846	4 CGAF000241	AF0000241 Gallus ga
7	17.4	87.0	4023	4 CHKNERB1	DI3719 Chicken mRN
8	17.4	87.0	113575	33 AL133354	AL133354 Homo sapi
9	17.4	87.0	113575	33 AL133354	AL121753 Homo sapi
10	17	85.0	153409	32 HSDJ61404	AL121752 Homo sapi
11	16.8	84.0	237415	32 HSDJ468N4	S77713 Ifi54-Inter
12	16.8	84.0	1742	12 S77710S2	247205 M.musculus
13	16.8	84.0	2551	12 MMRP2FCEN	M12223 Chicken tro
14	16.8	84.0	2571	4 CHKTR0S01	U43085 Mus musculu
15	16.8	84.0	3513	12 MM043085	AC008293 Drosophi1
16	16.8	84.0	55741	33 AC008293	AF000573 Homo sapi
17	16.8	84.0	55795	11 AF000573	AC008294 Drosophi1
18	16.8	84.0	56559	33 AC008294	AC016791 Mus muscu
19	16.8	84.0	76652	44 AC016791	AC011168 Homo sapi
20	16.8	84.0	89001	41 AC011168	AC007942 Mus muscu
21	16.8	84.0	123039	45 AC007942	AC009289 Homo sapi
22	16.8	84.0	141891	41 AC009289	AC013691 Homo sapi
23	16.8	84.0	169926	44 AC013691	G15848 human SFS C
24	16.4	82.0	2148	13 G15848	UB3438 Rickettsia
25	16.4	82.0	3188	2 RSU83438	X14298 Human mRNA
26	16.4	82.0	12446	9 HSDMDR	M18533 Homo sapien
27	16.4	82.0	13957	9 HSDMDR	M20250 Figure 1. N
28	16.4	82.0	13957	15 M20250	AC013624 Homo sapi
29	16.4	82.0	36083	43 AC013624	AC004059 Homo sapi
30	16.4	82.0	65201	45 AC004059	AC017270 Drosophi1
31	16.4	82.0	92053	45 AC017270	AC006061 Homo sapi
32	16.4	82.0	98056	40 AC006061	AC002981 Homo sapi
33	16.4	82.0	153568	11 AC002981	AC005832 Homo sapi
34	16.4	82.0	188948	40 AC005832	AC000016 Homo sapi
35	16.4	82.0	194000	33 AC000016	AC010547 Homo sapi
36	16.4	82.0	328651	42 AC010547	AC015434 Drosophi1
37	16	80.0	51648	43 AC015434	AC004352 Drosophi1
38	16	80.0	78643	35 AC004352	AC013362 Homo sapi
39	16	80.0	140692	44 AC013362	AL031684 Human DNA
40	16	80.0	143712	11 HS917N8	AC009589 Homo sapi
41	16	80.0	158301	42 AC009589	AC012274 Homo sapi
42	16	80.0	159469	42 AC012274	AC016801 Homo sapi
43	16	80.0	161598	45 AC016801	AL133163 Homo sapi
44	15.8	79.0	189565	32 CNS01D04	AF044224 Mus muscu
45	15.8	79.0	329	12 AF044224	A0028021 Rattus no

#### ALIGNMENTS

RESULT 1  
LOCUS AF023449 6110 bp mRNA  
DEFINITION Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)  
ACCESSION AF023449  
VERSION AF023449.1 GI:3169765



KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 6110)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a  
Down Syndrome Region and is Involved in the Development of the  
Nervous System  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6110)  
AUTHORS Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research  
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los  
Angeles, CA 90048-1869, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..6110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21 (trisomy 21)"  
/map="21q22, between HMG14 and MX1"  
/tissue\_type="brain"  
/dev\_stage="14 weeks, fetal"  
/clone="CHD2-42"  
/note="derived from alternately-spliced mRNA"  
<1..6110  
/gene="DSCAM"  
<1..5691  
/gene="DSCAM"  
/note="member of immunoglobulin superfamily; involved in  
nervous system development"  
/codon\_start=1  
/product="Down syndrome cell adhesion molecule"  
/protein\_id="AAC17966.1"  
/db\_xref="GI:3169766"  
/translation="VESEDLHSSLYPVNASLQVFASTTGTVPCPAGAPVYLRM  
YLATGEITDYPGIRHVPNGTLQIFPPPSSTSLHDTYICTAENSGKIRSDY  
HKAVALREPTYRVEDOKTRGNVAVAFKCIIPSSVAYITVSMEDKTVLSGSRFL  
ITSTGALYIKDVONEDGLNYRCITRHRVTGETROSNASALFVSDPANASPLDSD  
HRRAMAGORVELPCKALGHPEDYRMLKDMKPLESGRQKVTGLIENIRSDGS  
YVCEVSNRYGTAVIGRLYKOPRLAKTISPRKXSSGVSLSCSVTDEQELSMY  
RNEELNPKGNVRTGTHNENLIMDHVVKSDGAYOCFRKDKLSQDYVOYVLEGT  
PKLTISAFSEKVVSPAEFVSLMCNVKGPPLTITWTDLPILKGSHRISOMTSGN  
VSYLNISSQVHDGVRCTANNSAGVLYQARINVRPASIRPKNITATAAGRTY  
IHCVLIGYPIYSIKWYKNSNLPFNHROVAFENNGTLKLSVQKEVEDEEYCNLYO  
POLSTOSVHTVAKVPFIQPFEPFEPFISIGORVFIQCVVSGDLPITIMOKGRRPI  
GSLGVTIDNIDFTSLRISLIMHNGNCTCIARNEAAVEHQSOILIVAPREFVOP  
RDOGGYTKAVILKNSAEQYVPYIVKFEKAGVFOFOPIALNGRIQVLSNGLIK  
HYVEDSGITLCKVSDVGDVSKSMYLVKIVAMITSTPNTTLATQGGKRESCIAH  
GEPRIVRWKEDRIINPEMARLVSTKEGEEVISTQLPLTVREDSEFSCHLANS  
YGEDRGIIQLTVEPDPPEIEIKDVAKARTLRMTMGEDGNSPIGTGYIECKNDS  
WDSKQRTDVPOLNSATIIDIHPSSTYSIAMAKNRIGKESPNELTTADEAAGDS  
PPOVHLPEPISOSIRVTAKPKKHLONGIIRGOYIGREYSGNGNENIIVSDVGS  
DSFVYTLDNKFEQVGLVQACNRAGTGPSSQELITTTLEDVPSYPPENVAIASP  
ESISISMTLSKELNGILOGRFYITVIANLMODELGEIKNITTQPSLEDLGEKLTN  
YSIOVLAFTRAGDGVRSEQLFTRKEDVPEPAGVAAAASAVSMVPLPKLGI  
IRKATYVCSHPPTVISEFASPDSESYRIPLNSRNROSVWVAATVSGRNSSEII  
TVERLADAPARILITVKAEDSGYSCIANNMKSGDEIILNOVOPPDOPRLYTSKTS  
RSIFNSGSEFIITVKAEDSGYSCIANNMKSGDEIILNOVOPPDOPRLYTSKTS  
SSITLSWLPDGNSSIRGYILOYESNSFOMSGSPISPERSERLENTKCGMYEFT  
LTANONGCPRISEITIAKTLGKEPQSKDELFASINTVRANLIGNDGGCPRTS  
FTLEPRFGITVMTAORTSLSKSYLIDQEAFTWELQVNCNSAGCEKQANFTL  
NYGSGTIPLIKSVYONEGLTNEGKMLVTISCLVGLLLEFVLLVRRRROR  
LKRLRDKSLAEMLMSKNTSDTLNKOQTLRMHIDIPQAQLLIERPVTMETIDRS  
TLTLDADGEAAKOKSLIVTHTVYOSVATGCPLVDSADGATNPTRRAKAGP  
TANRRYASOMTLNRPPTISAHLLTDMRLPTPRAAGSVKDESDSYSPSDQTDAR  
SNMSTPSSASTYEELARAIEHAKMEQLPHAKFTTTECTISDTSEQLTAGNEETTD  
SLTSSTPSESGICRFTASPPKPDGGGVMNAVPAKAIQGVSTYICLHTLEWTFC"

BASE COUNT 1621 a 1658 c 1548 g 1283 t

ORIGIN

Query Match 100.0% Score 20; DB 11; Length 6110;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccagttctcaagagcagg 20  
|||||  
Db 4383 CCAGTTCCTCAAGAGACAGG 4402

RESULT 2  
AF023450  
LOCUS AF023450 6413 bp mRNA PRI 01-JUN-1998  
DEFINITION Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)  
mRNA, complete cds.  
ACCESSION AF023450  
VERSION AF023450.1 GI:3169767  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 6413)  
Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a  
Down Syndrome Region and is Involved in the Development of the  
Nervous System  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6413)  
AUTHORS Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,  
Lyons, G.E. and Korenberg, J.R.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research  
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los  
Angeles, CA 90048-1869, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..6413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21 (trisomy 21)"  
/map="21q22, between HMG14 and MX1"  
/tissue\_type="brain"  
/dev\_stage="14 weeks, fetal"  
/clone="CHD2-52"  
/note="derived from alternately-spliced mRNA"  
1..6413  
/gene="DSCAM"  
453..5168  
/gene="DSCAM"  
/note="member of immunoglobulin superfamily; involved in  
nervous system development"  
/codon\_start=1  
/product="Down syndrome cell adhesion molecule"  
/protein\_id="GI:3169768"  
/db\_xref="GI:3169768"  
/translation="MWILALSLFQSFANFSEDLHSSLYPVNASLQVFASTTGTLV  
PCPAGGIPVYLRMYLATGEITDYPGIRHVPNGTLQIFPPPSSTSLIHNTYYC  
TAENPSGIRSDOYHKAVALREPTYRVEDOKTRGNVAVAFKCIIPSSVAYITVSM  
EKDVSLVSGSRFLITSGALYIKDVONEDGLNYRCITRHRVTGETROSNASALFV  
SDPANASPLDSDHRRAMAGORVELPCKALGHPEDYRMLKDMKPLESGRQKVTGLI  
ENIRSDGSYVCEVSNRYGTAVIGRLYKOPRLAKTISPRKXSSGVSLSCSVTDEQEL  
CLIENTRPSDSGTYCEVSNRYGTAVIGRLYKOPRLAKTISPRKXSSGVSOVSL  
CSVGTDEQELSMYRNEELNPKGNVRTGTHNENLIMDHVVKSDGAYOCFRKDKL  
SAQDYVOYVLEDTGPKLTISAFSEKVVSPAEFVSLMCNVKGPPLTITWTDLPILK  
GSHISQMTSEGVAWSYLNSSQVHDGVRCTANNSAGVLYQARINVRPASIRPKNIT  
ATAAGRTYIHCVLIGYPIYSIKWYKNSNLPFNHROVAFENNGTLKLSVQKEVEDEE  
YCNLYOPOLSTOSVHTVAKVPFIQPFEPFEPFISIGORVFIQCVVSGDLPITIMOK  
EVDGEGTCANLYOPOLSTOSVHTVAKVPFIQPFEPFEPFISIGORVFIQCVVSGDL  
PITTMOKDGPRTGSGLTIDNIDFTSLRISLIMHNGNCTCIARNEAAVEHQSOILIV  
APREFVOPRDOGGYTKAVILKNSAEQYVPYIVKFEKAGVFOFOPIALNGRIQVLSN  
GLIKHYVEDSGITLCKVSDVGDVSKSMYLVKIVAMITSTPNTTLATQGGKRESCIAH  
GEPRIVRWKEDRIINPEMARLVSTKEGEEVISTQLPLTVREDSEFSCHLANSYGED  
RGIIQLTVEPDPPEIEIKDVAKARTLRMTMGEDGNSPIGTGYIECKNDSWDSKQRT  
DVPOLNSATIIDIHPSSTYSIAMAKNRIGKESPNELTTADEAAGDSPPOVHLPEPIS  
OSIRVTAKPKKHLONGIIRGOYIGREYSGNGNENIIVSDVGSDSFVYTLDNKFEQV  
GLVQACNRAGTGPSSQELITTTLEDVPSYPPENVAIASPESISISMTLSKELNGILO  
GRFYITVIANLMODELGEIKNITTQPSLEDLGEKLTNYSIOVLAFTRAGDGVRSEQL  
FTRKEDVPEPAGVAAAASAVSMVPLPKLGIIRKATYVCSHPPTVISEFASPDSESY  
RIPLNSRNROSVWVAATVSGRNSSEIITVERLADAPARILITVKAEDSGYSCIAN  
NMKSGDEIILNOVOPPDOPRLYTSKTSRSIFNSGSEFIITVKAEDSGYSCIANNMK  
SGDEIILNOVOPPDOPRLYTSKTSSSITLSWLPDGNSSIRGYILOYESNSFOMSGSP  
ISPERSERLENTKCGMYEFTLTANONGCPRISEITIAKTLGKEPQSKDELFASINTVR  
ANLIGNDGGCPRTSFTLEPRFGITVMTAORTSLSKSYLIDQEAFTWELQVNCNSAGCE  
KQANFTLNYGSGTIPLIKSVYONEGLTNEGKMLVTISCLVGLLLEFVLLVRRRROR  
LKRLRDKSLAEMLMSKNTSDTLNKOQTLRMHIDIPQAQLLIERPVTMETIDRS  
TLTLDADGEAAKOKSLIVTHTVYOSVATGCPLVDSADGATNPTRRAKAGPTANRRY  
ASOMTLNRPPTISAHLLTDMRLPTPRAAGSVKDESDSYSPSDQTDARSNMSTPSSAS  
TYEELARAIEHAKMEQLPHAKFTTTECTISDTSEQLTAGNEETDSLTSSTPSESGIC  
RFTASPPKPDGGGVMNAVPAKAIQGVSTYICLHTLEWTFC"



exon /evidence-not\_experimental  
11509..11621 exon prediction, score = 74% (0%)"  
/note="Xpound exon prediction, score = 74% (0%)"  
/evidence-not\_experimental  
exon /evidence-not\_experimental  
11535..11621  
/note="GRAIL, score = 58.000%, comment = good shadow"  
/evidence-not\_experimental  
repeat\_region 12393..12480  
/rpt\_family="L2"  
/evidence-not\_experimental  
repeat\_region complement(12492..12869)  
/rpt\_family="L1MB3"  
/evidence-not\_experimental  
repeat\_region 13831..14128  
/rpt\_family="AluSx"  
/evidence-not\_experimental  
repeat\_region 15597..15646  
/rpt\_family="L2"  
/evidence-not\_experimental  
exon complement(15672..15819)  
/note="WZEF, score = 53.3%"  
/evidence-not\_experimental  
repeat\_region 15717..15762  
/rpt\_family="PABL-A"  
/evidence-not\_experimental  
repeat\_region complement(17207..17470)  
/rpt\_family="AluSx"  
/evidence-not\_experimental  
exon complement(17433..17534)  
/note="GRAIL, score = 87.000%, comment = excellent"  
/evidence-not\_experimental  
repeat\_region complement(18789..19089)  
/rpt\_family="AluY"  
/evidence-not\_experimental  
repeat\_region complement(19490..20236)  
/rpt\_family="L1PA15"  
/evidence-not\_experimental  
repeat\_region complement(20244..26377)  
/rpt\_family="L1PA2"  
/evidence-not\_experimental  
exon complement(20451..22019)  
/note="GRAIL, score = 92.000%, comment = excellent"  
/evidence-not\_experimental  
exon complement(21819..21970)  
/note="WZEF, score = 65.5%"  
/evidence-not\_experimental  
exon complement(22035..22267)  
/note="GRAIL, score = 59.000%, comment = good"  
/evidence-not\_experimental  
exon complement(22537..22859)  
/note="GRAIL, score = 49.000%, comment = marginal"  
/evidence-not\_experimental  
exon complement(23004..24272)  
/note="GRAIL, score = 87.000%, comment = excellent"  
/evidence-not\_experimental  
repeat\_region complement(26405..27896)  
/rpt\_family="L1PA15"  
/evidence-not\_experimental  
exon complement(26651..26723)  
/note="GRAIL, score = 50.000%, comment = good"  
/evidence-not\_experimental  
exon complement(26651..26773)  
/note="WZEF, score = 52.1%"  
/evidence-not\_experimental  
repeat\_region 27887..29472  
/rpt\_family="L1P"  
/evidence-not\_experimental  
exon 28426..28580  
/note="WZEF, score = 63.6%"  
/evidence-not\_experimental  
repeat\_region 29473..29835  
/rpt\_family="THER1B"  
/evidence-not\_experimental

repeat\_region 29836..29969  
/rpt\_family="L1P"  
/evidence-not\_experimental  
repeat\_region complement(30121..30486)  
/rpt\_family="THER1C"  
/evidence-not\_experimental  
repeat\_region complement(30940..31299)  
/rpt\_family="L1MCI"  
/evidence-not\_experimental

Query Match 100.0%; Score 20; DB 11; Length 145861;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccaggtctcaagagcagg 0  
Db 56809 CCAGTCTCAAGAGCAGCAG 56828  
|||||  
|||||

RESULT 4  
AF064865/c AF064865 159424 bp DNA PRI 02-JUN-1998  
LOCUS Homo sapiens chromosome 21q22.3 PAC 58D10, complete sequence.  
DEFINITION AF064865  
ACCESSION AF064865.1 GI:3171160  
VERSION  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 159424)  
Bleichschmidt, K., Bagand, E., Hildmann, T., Nordsiek, G., Drescher, B.,  
Weber, J., Schattevoy, R., Yaspo, M.-L. and Rosenthal, A.  
Direct Submission  
TITLE Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
JOURNAL Location/Qualifiers  
FEATURES  
source 1..159424  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
complement(1224..1523)  
/rpt\_family="AluY"  
/evidence-not\_experimental  
1263..1386  
/note="GC score = 8.50 (124bp)"  
/note="Region: GC content"  
/evidence-not\_experimental  
1932..2047  
/note="WZEF, score = 92%"  
/evidence-not\_experimental  
3299..3348  
/note="Xpound exon prediction, score = 72% (0%)"  
/evidence-not\_experimental  
complement(3508..3614)  
/note="GenScan, score = -0.41%, comment = Terminal\_exon  
107 bp frame: 0 phase: 2"  
/evidence-not\_experimental  
3843..4124  
/note="GRAIL, score = 53.000%, comment = good shadow"  
/evidence-not\_experimental  
complement(3913..4018)  
/note="GenScan, score = 16.83%, comment = Initial\_exon 106  
bp frame: 1 phase: 1"  
/evidence-not\_experimental  
complement(3913..4089)  
/note="Xpound exon prediction, score = 94% (0%)"  
/evidence-not\_experimental  
complement(3913..4066)  
/note="WZEF, score = 72%"  
/evidence-not\_experimental  
complement(3913..4062)

exon	/note="GAIL, score = 93.000%, comment = excellent" /evidence-not_experimental complement(4787..4829) /note="Xpound exon prediction, score = 73% (0%)" /evidence-not_experimental complement(7468..7853) /note="Genscan, score = 35.77%, comment = Terminal_exon 386 bp frame: 0 phase: 2" /evidence-not_experimental complement(7529..7853) /note="GAIL, score = 86.000%, comment = excellent" /evidence-not_experimental complement(7546..7853) /note="Xpound exon prediction, score = 98% (0%)" /evidence-not_experimental complement(7563..7853) /note="MZF, score = 72.2%" /evidence-not_experimental complement(8472..8544) /note="GAIL, score = 49.000%, comment = marginal" /evidence-not_experimental complement(8772..8977) /note="GAIL, score = 60.000%, comment = good" /evidence-not_experimental complement(8780..8977) /note="Genscan, score = 1.30%, comment = Internal_exon 198 bp frame: 2 phase: 0" /evidence-not_experimental complement(9028..9216) /note="MZF, score = 96.1%" /evidence-not_experimental complement(9028..9216) /note="GAIL, score = 92.000%, comment = excellent" /evidence-not_experimental complement(9028..9216) /note="Genscan, score = 13.54%, comment = Internal_exon 189 bp frame: 1 phase: 0" /evidence-not_experimental complement(9642..9963) /rpt_family="MIR" /evidence-not_experimental 11055..11354 /rpt_family="AluY" /evidence-not_experimental 11606..11795 /rpt_family="MER63A" /evidence-not_experimental 12023..12324 /rpt_family="AluX" /evidence-not_experimental complement(12784..12882) /note="MZF, score = 96.1%" /evidence-not_experimental complement(12784..12882) /note="Genscan, score = 0.56%, comment = Internal_exon 99 bp frame: 1 phase: 0" /evidence-not_experimental 12814..12872 /note="GAIL, score = 59.000%, comment = good shadow" /evidence-not_experimental complement(13508..13527) /note="Xpound exon prediction, score = 67% (0%)" /evidence-not_experimental complement(14478..14641) /note="Genscan, score = 9.80%, comment = Internal_exon 164 bp frame: 0 phase: 2" /evidence-not_experimental complement(14478..14590) /note="Xpound exon prediction, score = 86% (0%)" /evidence-not_experimental complement(14478..14641) /note="GAIL, score = 94.000%, comment = excellent" /evidence-not_experimental	exon	complement(14478..14645) /note="MZF, score = 86.2%" /evidence-not_experimental complement(16046..16163) /note="Genscan, score = 6.30%, comment = Internal_exon 118 bp frame: 0 phase: 1" /evidence-not_experimental 16257..16330 /note="GAIL, score = 43.000%, comment = marginal shadow" /evidence-not_experimental 18126..18259 /rpt_family="MER5A" /evidence-not_experimental 19077..19876 /rpt_family="L2" /evidence-not_experimental 19945..20070 /rpt_family="LMC3" /evidence-not_experimental 20072..20405 /note="MZF, score = 84.2%" /evidence-not_experimental 20106..20254 /rpt_family="L2" /evidence-not_experimental complement(20145..20172) /note="Xpound exon prediction, score = 73% (0%)" /evidence-not_experimental complement(20491..20855) /rpt_family="THE1B" /evidence-not_experimental 20859..20899 /rpt_family="L2" /evidence-not_experimental 21282..21348 /rpt_family="L2" /evidence-not_experimental 21987..22074 /note="MZF, score = 85.6%" /evidence-not_experimental complement(22597..22750) /note="MZF, score = 99.6%" /evidence-not_experimental complement(22597..22750) /note="Xpound exon prediction, score = 87% (0%)" /evidence-not_experimental complement(22597..22750) /note="Genscan, score = 13.62%, comment = Internal_exon 154 bp frame: 1 phase: 1" /evidence-not_experimental complement(22597..22750) /note="GAIL, score = 100.000%, comment = excellent" /evidence-not_experimental 22995..23787 /rpt_family="L1M4" /evidence-not_experimental 23793..23859 /rpt_family="L1M8" /evidence-not_experimental complement(23861..24225) /rpt_family="THE1B" /evidence-not_experimental 24241..24582 /rpt_family="L1M8" /evidence-not_experimental 24395..24867 /rpt_family="L1M4" /evidence-not_experimental 24869..25346 /rpt_family="MER4A2" /evidence-not_experimental 25359..26535 /rpt_family="L1M4"
------	---	------	---

exon  
/evidence=not experimental  
26464..26634

Query Match 100.0%; Score 20; DB 11; Length 159424;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccagttctcaaggagcagg 20  
|||||  
Db 7849 CCAGTTCTCAAGGACGAGG 7830

RESULT 5  
LOCUS CHKPI05A 3654 bp mRNA VRT 28-Apr-1993  
DEFINITION Chicken p105 (NF-kappaB p50 precursor) mRNA, complete cds.  
ACCESSION M86930  
VERSION M86930.1 GI:212525  
KEYWORDS NF-kappaB p50 subunit precursor; p105 gene.  
SOURCE Gallus gallus (library: cDNA lambda zap) embryo fibroblasts cDNA to mRNA.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;  
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 3654)  
Capobianco,A.J., Chang,D.D., Mostafaei,G. and Gilmore,T.D.  
p105, the NF-kappa-B p50 precursor protein, is one of the cellular  
proteins complexed with the v-rel oncoprotein in transformed chicken  
splenic cells  
J. Virol. 66, 3758-3767 (1992)

JOURNAL MEDLINE  
FEATURES  
Source Location/Qualifiers  
1..3654  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/cell\_line="CEF, primary fibroblasts"  
/dev\_stage="embryo"  
/tissue\_type="fibroblasts"  
/tissue\_lib="cDNA lambda zap"  
1..217  
218..3172  
/gene="p105"  
218..3172  
/note="NF-kappaB p50 subunit precursor"  
/codon\_start=1  
/product="p105"  
/protein\_id="AA49000.1"  
/db\_xref="GI:212526"  
/translation="MAGEDPYIMGVSDPQMFAMDQLMGMSITFNNGYITSDLPRLTA  
DGPYLOITEPKRGKRFRRVYCEGSHGLPGASSEKNKSKYPOVKICNVGPAKVIY  
OLVNGKRYRHLHNSLVGKFCEDGVCYVAGPKDMVYGFANLGIHVTKKVEFTLET  
RMIDACKRGYNPGLIHPBELGVLQAECCGDRQLREIREIIRAOAAVOOTEMDSYVR  
LMFTAFIPDSNGGFTRRRLDPYISDAIYDSKAPNANSLKIYMDRTACGCVGEFTYLL  
CDKVOVDIQRFTBEDENGMMWEGFDFSTVHRQFAIYFKTIPKTRDNNITPAVY  
FVQLRRKSDLETSEKPLLYPEIKDKEVQRKKQLMPNFSIDYGGSGSGAGGGMG  
GGGAGSGGSESYSGYSAFGGMHFPGTTSNAGMKHLSNVKKRDEESSDQSK  
WTKHDKVETVEKNECRSGHNEKEEDASLCCEDEKPKCGCGDGLFEKAMQLAK  
KCNRLPYAVTGDVRLAVORHLTAVO..ONGNV..CHL..IHLHRELVKMLPEVMD  
KYNNTIIMNRNDLYOTPLHLAVITRKAQVVDLLKAGANVLLDRHGSNVHLAAAG  
DOKLITLLKHOKASMTIDLSNGELSAIHVVTANSLSCLLIIAGADVNOEOKS  
GRTALILEQENPVLACGLLEGDADVSTTYGTPPLHIAARGFTKPLAAVILKAAG  
ADPHVENLEPDLVEEDKDDDDGIVGTTPLDMAMNEVYDILNGKPIIAAAS  
EDLISGCPRLNESSKOOLYKLETTPDSKMSSTLAEKLGILINNAFOLEPSPST  
LIDNKKISGGTGOELIAFTQMDHTEALEVIOKALLSSORSHOEDRTIEAFPLSPST  
SFAKETGEIYNHKKFQPESTCDSEVSEFRLKSTFTYSDLSNKSITLSKMLTYGR  
KAOCAVYILTR"

BASE COUNT 1097 a 793 c 922 g 842 t  
ORIGIN

Query Match 87.0%; Score 17.4; DB 4; Length 3654;  
Best Local Similarity 94.7%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ccagttctcaaggagcagg 20  
|||||  
Db 2418 CAGTTCTCAAGGACGAGG 2436

RESULT 6  
LOCUS GGAF000241 3846 bp mRNA VRT 28-MAY-1997  
DEFINITION Gallus gallus nuclear factor NF-kB1 mRNA, complete cds.  
ACCESSION AF000241  
VERSION AF000241.1 GI:2130627  
KEYWORDS chicken.  
SOURCE Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;  
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 3846)  
Cabannes,E., Vives,M.-F. and Bedard,P.-A.  
Transcriptional and post-transcriptional regulation of  
KB-controlled genes by pp60v-src  
Oncogene (1997) in press  
2 (bases 1 to 3846)  
Bedard,P.-A., Cabannes,E. and Vives,M.-F.  
Direct Submission  
Submitted (18-Apr-1997) Biology, York University, 4700 Keele  
Street, North York, Ontario M3J 1P3, Canada

JOURNAL MEDLINE  
FEATURES  
Source Location/Qualifiers  
1..3846  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
5'UTR 1..160  
CDS 161..3178  
/codon\_start=1  
/product="nuclear factor NF-kB1"  
/protein\_id="AAB58343.1"  
/db\_xref="GI:2130628"  
/translation="MEATATATBARLNPCADPGILMAGEDPYIMGVSDPQMFAMDQLM  
GMSITFNNGYITSDLPRLTADGPLYLOITEPKRGKRFRRVYCEGSHGLPGASSEK  
NKSYPOVKICNVGPAKVIYOLVNGKRYRHLHNSLVGKFCEDGVCYVAGPKDMVY  
GFANLGIHVTKKVEFTLETMRIDACKRGYNPGLIHPBELGVLQAECCGDRQLRE  
REIRAOAAVOOTEMDSYVRLMFTAFIPDSNGGFTRRRLDPYISDAIYDSKAPNANSL  
KIVRMDRTACGCVGEFTYLLCDKVOVDIQRFTBEDENGMMWEGFDFSTVHRQ  
FAIVRTPKYRDNVITKPAVSFVQLRRKSDLETSEKPLLYPEIKDKEVQRKKQL  
MPNFSIDYGGSGAGGGMGCGGAGSGSESYSGYSAFGGMHFPGTTSNAGMK  
HLSNVTKKDEESSDQSKWTKHDKVETVEKNECRSGHNEKEEDASLCCEDEK  
NPKCGCGDGLFEKAMQLAKHCNRLPYAVTGDVRLAVORHLTAVODNDNVL  
HLSIHLHRELVKMLPEVMDKYNNTIIMNRNDLYOTPLHLAVITRKAQVVDLLKAG  
AVNLLDRHGSNVHLAAAGDOKLITLLKHOKASMTIDLSNGELSAIHVVTANS  
LSCLLIIAGADVNOEOKSGRTALILEQENPVLACGLLEGDADVSTTYGTPPLHIA  
ARGFTKPLAAVILKAAGADPHVENLEPDLVEEDKDDDDGIVGTTPLDMAMNEVYD  
ILNGKPIIAAASADPHVENLEPDLVEEDKDDDDGIVGTTPLDMAMNEVYDILNGKPI  
IAAASADPHVENLEPDLVEEDKDDDDGIVGTTPLDMAMNEVYDILNGKPIIAAAS  
EDLISGCPRLNESSKOOLYKLETTPDSKMSSTLAEKLGILINNAFOLEPSPST  
LIDNKKISGGTGOELIAFTQMDHTEALEVIOKALLSSORSHOEDRTIEAFPLSPST  
SFAKETGEIYNHKKFQPESTCDSEVSEFRLKSTFTYSDLSNKSITLSKMLTYGR  
KAOCAVYILTR"

BASE COUNT 1167 a 815 c 939 g 925 t  
ORIGIN

Query Match 87.0%; Score 17.4; DB 4; Length 3846;  
Best Local Similarity 94.7%; Pred. No. 38;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ccagttctcaaggagcagg 20  
|||||  
Db 2424 CAGTTCTCAAGGACGAGG 2442

RESULT 7  
LOCUS CHKNFKB1 4023 bp mRNA VRT 03-FEB-1999

```

DEFINITION   Chicken mRNA for NF-kB p50 subunit.
ACCESSION    D13719
VERSION       D13719.1 GI:222838
KEYWORDS      DNA binding protein; NF-kB p50; NF-kB/rel family.
SOURCE        Gallus gallus (library: lambda gtl1) spleen, cDNA to mRNA, clone
              PCR410.
ORGANISM      Gallus gallus
              Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
              Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE     1 (bases 1 to 4023)
AUTHORS       Ikeda,T.
TITLE          Direct Submission
              Submitted (19-NOV-1992) to the DDBJ/EMBL/GenBank databases. Toshio
              Ikeda, Institute for Virus Research, Kyoto University; Shogoin,
              Sakyo-Ku, Kyoto 606, Japan (E-mail: tikedae@virus.kyoto-u.ac.jp,
              tel:075-751-3990, Fax:075-751-3991)
              2 (bases 1 to 4023)
REFERENCE     Ikeda,T., Honjo,K., Hirota,Y. and Onodera,T.
              Isolation of the chicken NF-kappa B p65 subunit-encoding cDNA and
              characterization of its products
              Characterization of its products
              Gene 133 (2), 237-242 (1993)
JOURNAL        Submitted (19-NOV-1992) to DDBJ by:
MEDLINE        Toshio Ikeda
COMMENT        Institute for Virus Research
              Kyoto University
              Shogoin, Sakyo-ku
              Kyoto 606
              Japan
              Phone: 075-751-3990
              Fax: 075-751-3991
              E-mail: tikedae@virus.kyoto-u.ac.jp.
              Location/Qualifiers
                1..4023
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /clone_lib="lambda gtl1"
                /issue_type="spleen"
                404..3355
                /codon_start=1
                /product="NF-kB p50 subunit"
                /protein_id="BAA02872.1"
                /db_xref="GI:222838"
                /translation="MAGEDEPYIKGVSDPQMFAMDLQMLNSTIFNNITGYITSLDPLRTA
                DGPYIOTIEOPKQGRFRFRVYCEGSPHGGIPGASSSEKNKSYPOVVICNYVPAYIV
                QLVTKNGYVHLHSLVKGFCDEGCVTVNAGPKDMVVGANGLIHVTKKYELET
                RMIDACKKGYNPGFLVHPELGYDQEGCGDROLTEREERTITRAAYVQTKEDLSYVR
                LMTAFALPDSNGGFTRLDPLVISAIDAYSAPNASNLKIVRMDRTAGCVTGEETILL
                CDVVQDDIOIRFYEDENGAMGEGFDESPDVHROFAIVETPKYRDVNTTKPASF
                FQLRKRSDELTSEKPELYPEIKDKEVQRKRLMPNFSQYGGSGAGGCMFG
                GGGGAGSGCTSYSPGYSAFGCMHFRGTTKSNAGKHELNSITVKKDESSKODK
                WDTRKDVKVEIVKNECRTSGHNEEKEDASLCKDCKGNPKCCGCGCLPEKAMOLAK
                RHONALFEDVAVTGDVRLAVORHLAVODNDGDNVLSIHLHELKYLEVAPD
                MNYNNIINMRNDLYQTPHLAVITTKOAEVEDLLKAGANNLIDRIGNSVHLAAEG
                DDKILSLKHKRASSMIDLSENGEGLSAHMYVTANSLSCKLLIAGADVANAQEQKS
                GRTALHLAEQENIPLAGCLLEGDVADSTYDGTTPHLIAGRGFTKLAIVLAKAG
                ADPRVENFEPLFVEEDVDKDDDEGIVPQTPPLDMAANVEYDILNNAKPYLAASV
                EDLISQSPLELNESSKOQYKXILLETFRPKSMSTLAERKLGILNNAPRLSPSKT
                LDNYKISGTVQELINALTOMQHTETIEYOKALSSSOQIQYOENITAEPSISPT
                SFAKEETGELYNHRFQDPESITCDSGVETSRKLSFTYSDLSNKSSTITLSKMTLGIGQ
                ESSYOSSYIPN"
BASE COUNT    1188 a      866 c      997 g      972 t
ORIGIN
polya-signal  4011..4016
              1188 a      866 c      997 g      972 t
Query Match    87.0%; Score 17.4; DB 4; Length 4023;
Best Local Similarity 94.7%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 cagttcctaaaggacgag 20
      11111111111111111
Db 2604 CAGTTCCTCAAGCAGG 2622

```

```

RESULT      8
AL133354/c  113575 bp  DNA      HTG      03-DEC-1999
LOCUS        Homo sapiens chromosome 20 clone RP4-599111, *** SEQUENCING IN
DEFINITION   PROGRESS ***, in unordered pieces.
ACCESSION    AL133354
VERSION       AL133354.1 GI:6523628
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     1 (bases 1 to 113575)
AUTHORS       Burton,J
TITLE          Direct Submission
              Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: hunquery@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
JOURNAL        IMPORTANT: This sequence is unfinished and does not necessarily
              represent the correct sequence. Work on the sequence is in
              progress and the release of this data is based on the understanding
              that the sequence may change as work continues. The sequence may
              be contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc. Order of segments is not known; 800 n's separate
              segments. Contig_ID: 00010 Length: 2739bp
              Contig_ID: 00025 Length: 2967bp
              Contig_ID: 00028 Length: 2249bp
              Contig_ID: 00035 Length: 1470bp
              Contig_ID: 00036 Length: 1428bp
              Contig_ID: 00039 Length: 1821bp
              Contig_ID: 00055 Length: 1174bp
              Contig_ID: 00064 Length: 1448bp
              Contig_ID: 00065 Length: 5014bp
              Contig_ID: 00083 Length: 1029bp
              Contig_ID: 00096 Length: 1723bp
              Contig_ID: 00107 Length: 1027bp
              Contig_ID: 00124 Length: 1596bp
              Contig_ID: 00136 Length: 3878bp
              Contig_ID: 00141 Length: 1080bp
              Contig_ID: 00150 Length: 1820bp
              Contig_ID: 00193 Length: 1075bp
              Contig_ID: 00209 Length: 3504bp
              Contig_ID: 00224 Length: 2681bp
              Contig_ID: 00250 Length: 1771bp
              Contig_ID: 00253 Length: 1431bp
              Contig_ID: 00278 Length: 1129bp
              Contig_ID: 00292 Length: 1380bp
              Contig_ID: 00327 Length: 1536bp
              Contig_ID: 00329 Length: 1516bp
              Contig_ID: 00344 Length: 1814bp
              Contig_ID: 00354 Length: 4921bp
              Contig_ID: 00354 Length: 4921bp
              Contig_ID: 00367 Length: 1147bp
              Contig_ID: 00409 Length: 1143bp
              Contig_ID: 00418 Length: 3388bp
              Contig_ID: 00429 Length: 3151bp
              Contig_ID: 00434 Length: 1915bp
              Contig_ID: 00437 Length: 1068bp
              Contig_ID: 00480 Length: 2277bp
              Contig_ID: 00491 Length: 1303bp
              Contig_ID: 00500 Length: 1398bp
              Contig_ID: 00542 Length: 2072bp
              Contig_ID: 00564 Length: 2215bp
              Contig_ID: 00570 Length: 1152bp
              Contig_ID: 00571 Length: 1867bp
              Contig_ID: 00579 Length: 2258bp

```

```

FEATURES
source      1..113575
              Location/Qualifiers

```

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```







SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 355)  
AUTHORS Seizewsky, I., Hunt, E., Nguyen, M., Korn, B., Roehrdanz, B.,  
Lehrich, H., and Yaspo, M.L.H.  
TITLE An integrated transcript map for the whole human chromosome 21  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288678.  
Contact: Yaspo, M.-L.  
Max Planck Institut fuer Molekulare Genetik  
Imestrasse 73, D14195 Berlin-Dahlem, Germany  
Chr21.subseqec\_p112:  
Location/Qualifiers  
1..355  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="21q"  
/clone="MP112-218"  
/clone\_lib="Selected chromosome 21 cDNA library"  
BASE COUNT 83 a 83 c 90 g 99 t  
ORIGIN

alignment\_scores:  
Quality: 604.00 Length: 115  
Ratio: 5.345 Gaps: 0  
Percent Similarity: 98.261 Percent Identity: 98.261

alignment\_block:  
US-08-956-991-2 x AJ003472/rev ..

Align seq 1/1 to reverse of: AJ003472 from: 1 to: 355

1345 TleatgrrvalysAlagluaspsersgltyrrsercysllealaas 1362  
|||||  
353 ATTCCACGGTGAAGACAGAAAGACTCCGGCTATTACAGCTCATGCCAA 304  
|||||  
1362 nasnasntrpqlysersapgluilelleleuasnleuglnvalglnalp 1379  
|||||  
303 TAACAACTGGGGATCTGATGAATATTATTAACTTACAACTACAACTTC 254  
|||||  
1379 rorproaspqlnproargleuthrvalserlsthtrhserSerSerle 1395  
|||||  
253 CACCAGATCAGCCTCGCCTATAGCTCCAGACCACCTCTTCTCCATC 204  
|||||  
1395 ThrleusertpleuproglyAspasnglylserSerleleargglyty 1412  
|||||  
203 ACCCTTTTGGCTCCTGGAGACACAGGGGGCACTTATCAGAGATA 154  
|||||  
1412 rllleuglntrsergluaspsasnsergluglntrpqlyserphepro 1429  
|||||  
153 CATACTGAGTACTCCGAGACAATAGTACAGAGTGGGAGATTTCCTCA 104  
|||||  
1429 leserprosergluarqserlyrargleuglnuasnleulyscylgylth 1445  
|||||  
103 TCAGCCCGAGCGAAGCTTCTATCGCTAGCAAAATCTCAAAATGGGGCT 54  
|||||  
1446 TrpTyrLysPheThrLeuThrAlaGlnasnGlyValGlyProGly 1460  
|||||  
53 TGGTATAGTTCACACTACAGCCCAAAATGAGTGGGCCAGGG 9

seq\_name: qb\_gss1:FR0002941

seq\_documentation\_block:  
LOCUS FR0002941 619 bp DNA GSS 27-FEB-1997  
DEFINITION F.rubripes GSS sequence, clone 015616a10, genomic survey sequence.  
ACCESSION Z86724  
VERSION Z86724.1 GI:1883636  
KEYWORDS GSS: genome survey sequence.  
SOURCE Fugu rubripes.  
ORGANISM Fugu rubripes

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.  
REFERENCE 1 (bases 1 to 619)  
AUTHORS Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrania, Y.,  
Williams, G., and Brenner, S.  
TITLE Direct Submission  
JOURNAL Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk  
COMMENT Vector: pBluescript II KS  
V\_type: phagemid  
PRIMER: KS  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

FEATURES  
source Location/Qualifiers  
1..619  
/organism="Fugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 015616"  
/clone="015616a10"  
BASE COUNT 133 a 169 c 141 g 156 t 20 others  
ORIGIN

alignment\_scores:  
Quality: 576.00 Length: 181  
Ratio: 4.204 Gaps: 3  
Percent Similarity: 75.691 Percent Identity: 66.298

alignment\_block:  
US-08-956-991-2 x FR0002941 ..

Align seq 1/1 to: FR0002941 from: 1 to: 619

22 SerSerLeuTyrPheValAlaSerLeuGlnGluValAlaPheAla 38  
|||||  
3 TCACGCTTATTATTGTTCATGCATCTGTGCAAGAGAGTGTTCCTCAT 52  
|||||  
38 TrpThrThr GlyThrLeuValProCysProAlaAlaGlyLeuProProVal 54  
|||||  
53 CACCACAGGAGTACGTTGGTCCCTGTCCGCTGCAGAGTGGCCCCCTTC 102  
|||||  
55 ThrLeuArgTrpTyrLeuValaThrGlyGluGluIleTyrAspValProG 71  
|||||  
103 TCATTCAGGTGATCTCCCTACTGACGAGAGATCTTCAGATGTCGAC 152  
|||||  
71 YlleatghisvalahisproasnGlyThrleuglnllepheprophepro 88  
|||||  
153 NATCCGACAGGTGCACCTCAATGGCACCCCTTCAGATCTTCACATCC 202  
|||||  
88 roserSerPheSerThrLeuIleHisaspasnThrTyrTyrCysThrAla 104  
|||||  
203 CGTCATATGTCANCAACATCATCATGACACACACTACTACTGCACAGC 252  
|||||  
105 GluAsnProSerGlyLyslleargserGlnaspValHisIleLys.... 119  
|||||  
253 GAGAACCCCTCAGGAGAAATCAGACCCNACATGCTTACATTAAAGCTGG 302  
|||||  
119 ..... 119

303 TGGCGTCCGGGGNNTACCTGCNCTCCANATNTTCCAGACTGACGTG 352  
|||||  
120 .....Ala..ValleuArgLupr 126  
|||||  
353 TTCCGANGCTNNCTGTGTCTCTCCCATNNCCAGTCTCTACNAGACCT 402  
|||||  
126 YrThrValArgValGluaspGlnLysThrMetArgGlyAsnValAlaVal 142  
|||||  
403 ACACGGTCCGGGAGAGATCAAAAAGATATGCGAGGAGCTGCACAGTC 452  
|||||  
143 PheLysCysIleIleProSerSerValGluAlaTyrIleThrValAlaIse 159  
|||||

```

453 TTCACGTGATTATCCCTGCTAGTGAGACATATATCATCTGTGTGTC 502
159 rTTPGtLysAspThrValSerLeuValSerGlySerArg 172
||||| |||||||:|||||:|:|:| |||:|:|
503 CTGGGANAAGACACCATGTCTCAATTATCTGAAAGTAG 542

seq_name: qb_est2:F13426

seq_documentation_block:
LOCUS F13426 310 bp mRNA EST 15-MAR-1995
DEFINITION HSC2X021 normalized infant brain cDNA Homo sapiens cDNA clone
c-2xa02, mRNA sequence.
ACCESSION F13426
VERSION F13426.1 GI:710043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 310)
Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C.,
Lorenzo,F., Mitchell,H., Marriage-Samson,R., Jumeau,M.N., Lamy,B.,
Sebastiani-Kabakchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
JOURNAL
MEDLINE
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: y2c-2xa02
Insert Length: 1500 Std Error: 0.00
Seq primer: (-21)M13-universal
High quality sequence stop: 335.
FEATURES
Source
1..310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type-total
brain; total mRNA was oligo-(dty) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaers, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S in press"
BASE COUNT 79 a 73 c 89 g 66 t 3 others
ORIGIN

alignment_scores:
Quality: 446.00 Length: 103
Ratio: 4.505 Gaps: 0
Percent similarity: 96.117 Percent identity: 95.146

alignment_block:
US-08-956-991-2 x F13426 ..
Alion seq 1/1 to: F13426 from: 1 to: 310
1610 PheValLeuLeuValValArgArgArgArgGluGluArgLeuLy 1626
||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

---

```

2 TTGTCTCTCTCTGTTGTTCTCCGAGAGCGCGGAGAG. AGCTTAA 50
1626 sATgLeuArgAspAlaLysSerLeuAlaGluLeuMetSerLysAsn 1643
||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GAGCGCTCGGAGATGCAAAAGATTACGAAATNTCATGAGTAACATA 100
1643 hArArgThrSerAspThrLeuSerLysGlnGlnGlnThrLeuArgMetHis 1659
||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 CCCGACTTCAGTACGTTAAGCAAGCACAGCAGCCTCGCAATGCAC 150
1660 lLeAspIleProArgAlaGlnLeuLeuIleGluGluArgAspThrMetG1 1676
||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ATCGCATACCCAGGCGCTCAGCTTTGATTGAAGAGAGACACCATGGA 200
1676 uThrIleAspAspArgSerThrValLeuLeuThrAspAlaAspPheGlyG 1693
||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 GACCATTTGATGATCGCTCCACGGCTCTNTTGACGAGCTGACTTTGAG 250
1693 LuAlaAlaLysGlnLysSerLeuThrValThrHisThrValHisTyrGln 1709
||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 GGGCAGCTTAAGCAGAAAGTCCTGACGGTCACACGGTCCATTACCAA 300
1710 SerValSer 1712
||||| |||||||
301 TCGGTGTCT 309

seq_name: qb_est1:245894

seq_documentation_block:
LOCUS 245894 294 bp mRNA EST 14-NOV-1994
DEFINITION HSC2MH041 normalized infant brain cDNA Homo sapiens cDNA clone
c-2wh04, mRNA sequence.
ACCESSION 245894
VERSION 245894.1 GI:575128
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 294)
Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houllgate,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Marriage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
JOURNAL
MEDLINE
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: yle-2wh04
Seq primer: (-21)M13-universal.
FEATURES
Source
1..294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-2wh04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type-total
brain; total mRNA was oligo-(dty) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaers, Psychiatry

```



```

REFERENCE      Pterygota: Neoptera: Endopterygota: Diptera: Brachycera:
AUTHORS        Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
TITLE          1 (bases 1 to 646)
JOURNAL        Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
                Brokstein, P., Lewis, S. and Rubin, G.M.
                BDGP/HMI Drosophila EST Project
                Unpublished (1997)
COMMENT        On Jan 19, 1998 this sequence version replaced gi:2151874.
                Contact: Harvey, D.
                G. M. Rubin-Molecular and Cell Biology
                University of California Berkeley
                539 LSA, Berkeley, CA 94720-3200, USA
                Fax: 510 643 9947
                Email: http://www.fruitfly.org/EST\_estefruitfly.berkeley.edu
                Plate: 61 row: C column: 10
                High quality sequence stop: 481.
FEATURES
  source        1..646
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="GH06134"
                /clone_1fb="GH Drosophila melanogaster head port2"
                /sex="male and female"
                /dev_stage="adult"
                /lab_host="DHS - alpha"
                /note="Organ: head; Vector: pORT2; Site_1: EcoRI; Site_2:
                XhoI; Sized fractionated cDNAs were directly ligated into
                pORT2. Plasmid cDNA library."
BASE COUNT     165 a 175 c 166 g 140 t
ORIGIN
alignment_scores:
  Quality:      411.00      Length:      195
  Ratio:        2.936      Gaps:      1
  Percent Similarity: 71.795  Percent Identity: 42.564
alignment_block:
US-08-956-991-2 x A1106884 ..
Align seq 1/1 to: A1106884 from: 1 to: 646
1033 PheGlnPheAsnIleIleSerValIleThrSergIuAspSergIuValIy 1049
      :::::::::::::::::::: ||| ::::||| ::::
4 TAATACCTTACTCCGTTCCGAGACGCCGATGAGCATGCGACCT 53
1049 rThrlEuAspAsnLeuAsnIySpherhGlnTyrgIyLeuValVa..cIlna 1066
      |||::::::::::||| |||||::::::::::|||
54 ACTACTCAGCGAGCTGGCTAGTTGCCAGATACACCGTGCTGCTGACAG 103
1056 IeCyAsnArGAlaGlyThrlGlyProSerSergIuIleIleThrThr 1082
      |||::::::::::||| |||||::::::::::|||
104 CCTTCATCAATGATGGCGCCAGGACCTCTTCGGAAACCCAGCGTGCACG 153
1083 ThrLeuGluAspValProSerTyTyrProGluAsnValGlnaIleal 1099
      |||::::::::::||| |||||::::::::::|||
154 ACTATGAGAAGATG1CCAGTCGACACACCGAGAGACTTGCTGTGGCCG 203
1099 aThrSerProGluSerIleSerIleSerTTPSerThrlLeuSerLyGluA 1116
      | ||| ::::|::::::::::|||::::::::::
204 CTATTCCTCCCAATGCTTCACGAGTGCTCGGCAACCCACCCCAATTAC 253
1116 IeLeuAsnGlyIleLeuGlnGlyPheArGValIleTyTyrPalaaSleu 1132
      |||||::::::::::|||::::::::::|||::::::::::
254 ACACCAATGCGCTGCTGCAGGATACCAAGCTGATTCGAGCCCATCATC 303
3 MetAsp.....GlyGluLeuGlyGluIleLyAsnIleThrThrThreI 1147
      ||| :::: |||::::::::::|||::::::::::
7ATGACATTACAGCCAGCAGACGAAAGTGAAGTCGAGGAAAGACGACG 353
+oserLeuGluLeuAspGlyLeuGluLyTyTyrThrAsnTySerIleG 1164
      ::::::||| |||||::::::::::|||::::::::::
+CAATGGTGTGACAGGAGCTTTCGAAATATACCAACCACTACAGATTC 403

```

```

1164 lnualeuAlaPheThrArgAlaGlyAspGlyValArgSerGluGlnIle 1180
|||||
404 AGGTTTGCCACACAGCGATGGCGATGGAGTGATGTCGAACCCATTG 453
1181 PheThrArgThrGlySerGluAspValProGlyProProAlaGlyValLysAl 1197
|||||
454 TTTTGCCACACAGCGATGGCTAGAGCTCCGCGAGATATCAAGT 503
1197 aAlaAlaAlaSerAlaSerMetValPheValSerTrpLeuProLeuL 1214
|||||
504 TGTGTCCAGCTCATCTCAATCACTGATATCTCTGTGTTCCACCCAAATG 553
1214 yslleuAsnGlyIleIleArgly) [rYrThValPhe 1225
554 AACCGAACGGTGATGATCCACAGATGAGTCTTTC 588
seq_name: gb_est38:AW048129

seq_documentation_block:
LOCUS AW048129 402 bp mRNA EST 18-SEP-1999
DEFINITION UT-M-BH1-alq-f-04-0-UT.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
VERSION UT-M-BH1-alq-f-04-0-UT.3, mRNA sequence.
ACCESSION AM048129
VERSION AM048129.1 GI:5908658
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 402)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247317.
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301.443.1706
Fax: 301.443.9890
Email: mestr@nimh.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized brain stems library CNIH CNA Library Preparation: M.B.
Soares lab clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
The following repetitive elements were found in this cDNA sequence:
1-28, >AT-rich; low-complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES
Source
Location/Qualifiers
1..402
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="UT-M-BH1-alq-f-04-0-UT"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S2 library is a subtracted library derived from
NIH_BMAP_M_S1, which in turn is a subtracted library
derived from a mixture of normalized libraries from ten
regions of the mouse brain (cerebellum, brain stems,

```

```

IAC_LIB=NIH_BMAP_M_S2
TAG_TISSUE=brain-stems
TAG_SEQ=TCATG"

```

```
alignment_scores:
  Quality: 408.50
  Ratio: 3.782
  Percent Similarity: 82.443
  Length: 131
  Gaps: 1
  Percent Identity: 56.489
```

Align seg 1/1 to reverse of: AW048129 from: 1 to: 402

BASE COUNT	203 a	260 c	186 g	137 f
ORIGIN				

alignment_scores:		
Quality:	408.50	Length: 263
Ratio:	2.220	Gaps: 5
Percent Similarity:	69.962	Percent Identity: 37.643

Align seg 1/1 to: AA697364 from: 1 to: 786

1152 LeuAspGlyLeuGluLysTyrThrAsnTyrSerIleGlnValLeuAlaPh 1168  
||| |||||:::|||||||:::|||||||  
9 CTCATGGCTTAAGAAGTACACCACTACACATGCAGTCTCGGCC.. 56

1168 eThrArgAlaGlyAsp.GlyValArgSerGluGlnIlePheThrArgThr 1184

57 .ACAACAGCCGGAGTGAAGCGTCGCGAGCGTTCCCATCCACTGCCAACC 105  
1185 LysGIuaspValPProGLyProProlagIvalLysAlaIalalase 1201

106 GAACCGATGTTCCCGAAGCACCCACTGATGTTAAAGGCTCTTGTTATGGG 155

1201 rAlaSerMetValPheValSerIrrpleuProProLeuLysLeuAnsnglyI 1218  
:  
:: :::::::::::::::::::: ||||| :: |  
156 TAACGCCGCTATCCTGTATCTGGCGCCACCAGCACAGCCAACGGCA 205

1218 leIeaRgLyStYrThrValRheCysSerHisProTyrProThrValIle 1234

1235 SerGluupheGluAAserPro AsnSerpheSerTurratIleProIenI 1251  
206 TTATCACCACGATACACCGGTGTACTCCAAAGCCGAAGCGTGAAGACTGAGAC 255

256 AAAGACCCAGAGGTTCCCCACTACAGATGAGTTTCGAGGCCACCGAAC 305

```

1251 euseRrAqasnaRGIntyrSeValTrpaValAlaValThrSerAla 1267
||:::||::| ||::| ||||::|| |:::
306 TCCAGACAGACAACCCTTCAACCTCCCCCCTAACCTCACACGCGATTM 355

```

1268 GlyArgGlyAsnSerSerGluIleIleThrValGluProLeuAlaLysAl 1284

356 GCGGAACGCACTCAGTCTAAGAGCATAGTGGCCATGCCAGCAGCAGGTG 404

```

1204 a p r o l a a r g i l e l e u n r p h e s e r g y n r v a l t h r t h r p r o t i p m e l t 1300
      |||||.:||| :.:|||.: ||| |||.: :.: :
405 . c c c g c c a g a t c g c t c c t t c g a c g a c a c c t t c a c t g c c a c c t t c a a g 454

```

```

1301 ysaapilvalleuProcysLysalavalajlyaspproserProalaval 1317
      ::::::::::::::::::::
455 AGACGCCAGAGATCGCTGCGCTGGAGAGCCGCCAACCAGAGATC 504
      ::::::::::::::::::::

1318 LysTrpMetLysaspSerasnGlyThrProSerLeuValThrIleaspG 1334
      ::::::::::::::::::::
505 ACATGG.....AAGATCAAGGCGGTGAA.....TTCAGTCCCAACGA 542
      ::::::::::::::::::::

1334 YargArgSerIlePheSerasnGlySerPheIleIleArgThrValLysa 1351
      ::::::::::::::::::::
543 TCGCATCGCGCTCTCCCGCAGCGATCGCTGATCAATCGGTAATC 592
      ::::::::::::::::::::

1351 IagLysaspSerGlyTyrTyrSerCysIlealaasnasnasnTrpGlySer 1367
      ::::::::::::::::::::
593 GCCAGCATGCCGCGAGACTCTCGCCACGCCGAGAACTGATTCGTAAG 642
      ::::::::::::::::::::

1368 AspGluIleIleLeuasnLeuGlnValGlnValProProaspGlnProAr 1384
      ::::::::::::::::::::
643 GACTTCATTCACGCCACAAACTGATTTGTTCTGGACACACCAATCGCCCA 692
      ::::::::::::::::::::

1384 gLeuThrValSerIleThrThrSerSerIleThrLeuSerTrpLeup 1401
      ::::::::::::::::::::
693 CGTCACCCCTCTCGGCACACCAATGATGCCCTGACTGTAAGTTGAAGC 742
      ::::::::::::::::::::

1401 rGlyaspasnGlyGlySerSerIleArgGlyTyr 1412
      ::::::::::::::::::::
743 CCCATGAGGAGACACTGCTTCTGTCATGATGATAC 777
      ::::::::::::::::::::

seq_name: gb_gss3:B07180

seq_documentation_block:
LOCUS      B07180          465 bp          DNA          GSS          13-JUN-1996
DEFINITION CSRL72-h3-u CSRL flow sorted Chromosome 11 specific cosmid Homo
ACCESSION   B07180
VERSION     B07180.1  GI:41416458
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 465)
AUTHORS     Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S.,
            Harris,J., Deford,J., McFarland,J., Burzinski,K., Khan,M.,
            Kupfer,K. and Garner,H.R.
            Genomic Sequence Sampled Map of Chromosome 11
            Unpublished (1996)
            Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gervans@utsu.smed.edu, shane@mcdermott.smed.edu
TITLE       PCR Primers
JOURNAL     FORWARD CTGTCAGATGAAGCACTC
            BACKWARD: TTCTTTTAATCCCGAC
            Seq primer: 17
            Class: cosmid ends
            High quality sequence stop: 465.
FEATURES
            source
            1..465
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSRL72-h3"
            /clone_lib="CSRL flow sorted Chromosome 11 specific:
            cosmid"
            /sex="female"
            /cell_type="chimeric hamster somatic cell hybrid"
            /note="Vector: scos-1; Human Chromosome 11 specific cosmid
            library prepared from flow sorted human Chromosome 11

```

```

BASE COUNT      77 a      107 c      145 g      112 t      24 others
ORIGIN

alignment_scores:
      Quality: 379.50      Length: 98
      Ratio: 4.170      Gaps: 2
Percent Similarity: 92.857      Percent Identity: 78.571

alignment_block:
US-08-956-991-2 x B07180/rev
Align seg 1/1 to reverse of: B07180 from: 1 to: 465

1795 AsparGAlaArgSerSerMetValSerThrGluSerAlaSerThrTy 1811
      ::::::::::::::::::::
429 GACAAAGGAGAACAGCATGTGTCCATGAGAGTCTTTCACCTA 380
      ::::::::::::::::::::

1811 rGluGluLeuAlaArgAlaTyrGluHisAlaLysMetGluGluLeuA 1828
      ::::::::::::::::::::
379 CGAGAGCTGCGCCCGCGCNTATGACATGCCAAGCTGAGAGCAGCTCC 330
      ::::::::::::::::::::

1828 rGHisAlaLysPheThrIleThrGluCysPheIleSerAspThr.SerSe 1844
      ::::::::::::::::::::
329 AGCAGCGCCAACTTGAGATCACCGAGTCTTCATCTCTGACAGTTTCNTC 280
      ::::::::::::::::::::

1844 rGluGluLeuThrAlaGlyThrAsnGluTyrThrAspSerLeuThrSer. 1860
      ::::::::::::::::::::
279 TGACCGAG.TGACCAAGGACACCAAGAGAACGCCGACACATGACATCA 230
      ::::::::::::::::::::

1861 ..SerThrProSerGluSerGlyIleCysArgPheThrAlaSerProPro 1876
      ::::::::::::::::::::
229 TGAGCAGACACNTCAGAGACNTGCATCTGCCGNTTTACCGCTCACACCC 180
      ::::::::::::::::::::

1877 LysProGlnaspGlyGlyArgValMetasnMetAlaValPro 1890
      ::::::::::::::::::::
179 AAGCCCGAGATGCGGACCGGCCAAACCTGCTGTGCC 138
      ::::::::::::::::::::

seq_name: gb_est27:AI404957

seq_documentation_block:
LOCUS      AI404957          569 bp          mRNA          EST          08-FEB-1999
DEFINITION GH24836.5prime GH Drosophila melanogaster head P012 Drosophila
ACCESSION   AI404957
VERSION     AI404957.1  GI:4248044
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 569)
AUTHORS     Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Broksstein,P., Lewis,S. and Rubin,G.M.
            BDGP/HMI Drosophila EST Project
            Unpublished (1997)
            On Apr 21, 1998 this sequence version replaced gi:3073067.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            based upon the presence of vector sequence at both ends, this
            sequence has been determined to be the complete cDNA insert.
            Plate: 248 row: C column: 12
            High quality sequence stop: 560.
FEATURES
            source
            1..569
            /organism="Drosophila melanogaster"

```

```

/db_xref="taxon:7227"
/clone="GH24836"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site:1: EcoRI; Site:2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT      145 a      150 c      154 g      120 t
ORIGIN

```

```

alignment_scores:
  Quality: 355.50      Length: 185
  Ratio: 2.799        Gaps: 4
  Percent Similarity: 68.649      Percent Identity: 40.541

```

```

alignment_block:
  US-08-956-991-2 x A1404957 ..

```

```

Align seg 1/1 to: A1404957 from: 1 to: 569

```

```

1145 ThrhrclnProSerLeuGluLeuAspGlyLeuGlyLysThrAsnTy 1161
      ||||| ..... ||| |||||:|||||
7  ACACACACACTCACCATGCTGTCACAGACTTCGCAATATACCACTA 56
1161 rSerIleGlnValLeuAlaPheThrArgAlaGlyAspGlyValArgSerg 1178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57  CAGATTACAGTTTGGCCACACAGCGTATGGCGATGGAGTGTATCGA 106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1178 LucInIlePheThrArgThrLysGluAspValProGlyProAlaGly 1194
      :: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
107  AGCCATGTTTTCACACAGCAAGAGATGTGCTGAGCTCGCGCAGAT 156
1195 ValLysAlaAlaAlaSerAlaSerMetValPheValSerPheLeuP 1211
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
157  ATCAAGTTGTGTCACACTATCTCAATCAGTATATCTTGTGTTGCC 206
1211 oProLeuLysLeuAsnGlyIleIleArgLysThrValPheCysSerH 1228
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
207  ACCCAATGACCGACGCGTGTATCACAAGTACAGTCTT ..... 246
1228 IsProTyrProThrValIle ..... SerGluPhe 1237
      ||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
247  ....TACACTCGCTAGTGAACGTCGCGAGCACTGAACAATGAGAA 291
1238 GluAlaSerProAspSerPheSer...TyrArgIleProAsnLeuSerAr 1253
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
292  CGTAGTCTGCATCACAGCAGCGGTATTACGAAGCCAAAGAGTGCATCC 341
1253 gAsnATGGINTySerValITrpValAlaValAlaThrSerAlaGlyArgG 1270
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
342  GCATATGAGATACAGTCTGCTGACGCGCAGCACTGATGCGGAGG 391
1270 LysAsnSerGluIleIleThrValGluProLeuAlaLysAlaProAla 1286
      ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
392  GCAGAGTTCCCGTGTTCCTTCACAATACCAAGATCGCATCGCGGG 441
1287 ArgIleLeuThrPheSerGlyThrValIThrThrProTrpMetLysAsp 1303
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
442  AGGATCATATCTCTGCGTGCAGTGTGACAGCCTTGCGATCCACCGT 491
1303 eValLeuProCysLysAlaValAlaGlyAspProSerProAlaValITr 1320
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
492  TACTTACCTTCACGCGGTGCGCAAAACCAACGCC ..... GAGTGT 535
1320 eLys 1321
      ::|||
536  TCAA 540

```

```

seq_name: gb_gss1:FR0021997

```

```

seq documentation block:
LOCUS      FR0021997      610 bp      DNA      GSS      10-DEC-1997
DEFINITION F. rubripes GSS sequence, clone 070J16AB10, genomic survey sequence.
ACCESSION  AL014868
VERSION    AL014868.1 GI:2681236
KEYWORDS   GSS: genome survey sequence.
SOURCE     Fugu rubripes.
ORGANISM   Fugu rubripes.
REFERENCE  Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
AUTHORS    Williams,G. and Brenner,S.
TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
          Centre Hinxton, Cambridge, CB10 1SB. Email: biolhel@hmp.mrc.ac.uk
COMMENT    Vector: pBluescript II KS
          V.type: phagemid
          PRIMER: KS
DESCR      One pass dye-terminator sequencing of cosmid cloned genomic
          sequence.
FEATURES
  source          location/Qualifiers
                  1..610
                  /organism="Fugu rubripes"
                  /db_xref="taxon:31033"
                  /clone_lib="cosmid 070J16"
                  /clone="070J16AB10"
BASE COUNT      158 a      133 c      150 g      147 t      22 others
ORIGIN

```

```

alignment_scores:

```

```

  Quality: 346.50      Length: 187
  Ratio: 3.013        Gaps: 7
  Percent Similarity: 61.497      Percent Identity: 43.850

```

```

alignment_block:

```

```

  US-08-956-991-2 x FR0021997/rev ..

```

```

Align seg 1/1 to reverse of: FR0021997 from: 1 to: 610

```

```

586 SerGlnSerValHisValThrValLysValProPheIleGlnProPh 602
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
555  TCTTCTCTTCCACTTA ..... CTGTCACCTTCATCCACGCTT 515
602 eGluPheProArgPheSerIleGlyGlnArgValPheIleProCysVal 619
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
514  NGATATCCNTTNTACNTCAGTCGGGAGAGATCATTTACATCGCGTGG 465
619 aValSerGlyAspLeuProIleThrIleThrTrpGlnLysAspGlyAr 635
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
464  TGTATNTGGGACATGCCATCGCATCACTGCGCAAAAGATGGCCA 415
635 gProIle...ProGlySer...Leu.GlyValThrIleAspAsnIleAs 649
      ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
414  GCCATTGTGCCAGNCTCCGCTGCGCTGCGCAATAGAGACCAAGAG 365
650 PheThrSerSerLeuArgIleSerAsnLeuSerLeuMetHisAsnGly 666
      ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
364  TTCATGAGTTCTNTGCAAAATCACCAAGTGACGTTAAAGCAACAGAA 315
666 nTyThrCysIleAlaArgAsnGlnAlaAlaValGlnHisGlnSerg 683
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314  CTACAGTCGATCGCCACGCAACACAGCAGCTGTACGTTGGAGCGAG 265
683 InLeuIleVal..Arg ..... 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264  AGCTGATGTTTACAGGTGAGGAGGATTTACACCAAGAGTGCACGANT 215
687 ..... 687

```

```

214 TAGTCGGCTCAAAATTGTGATCTACCCAGAGAGACCCNATNNATAAG 165
688 .....ValProPolysph 692
164 TATTTCCTAGCTTATACATCTGCTGTTTTCAGTCCGCCGCTT 115
692 eValValGlnProArGAspGlnAspGly1LeTyrGlyLysAlaVal1LeI 709
114 TGTACTGCAGCCCAACATCAAGACTGATTATGCAAACTGGAGTGTG 65
709 euAsnCysSerAlaGluGlyTyrProValProThr1LeVal1TrpLysPhe 725
64 TCAACTGTTNTGCGAGGGGTACCTCCCCCAAGATCATGTGAAAC 15
726 SerLys 727
14 GCCAA 9
seq_name: gb_est28:A1542894
seq_documentation_block:
LOCUS A1542894 696 bp mRNA EST 22-MAR-1999
DEFINITION SD09407.5:prime SD Drosophila melanogaster Schneider L2 cell culture
sequence.
pot2 Drosophila melanogaster cDNA clone SD09407 5prime, mRNA
ACCESSION A1542894
VERSION A1542894.1 GI:4460267
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 696)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HM1 Drosophila EST project
Unpublished (1997)
On Mar 16, 1998 this sequence version replaced gi:2961848.
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST\_estfruitfly.berkeley.edu
Plate: 94. row: A. column: 7
High quality sequence stop: 558.
FEATURES
Source
1. 696
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD09407"
/culture_pot2="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pot2; site_1: EcoRI; site_2: XhoI. Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT 183 a 197 c 183 g 133 t
ORIGIN
alignment_scores:
Quality: 345.50 Length: 230
Ratio: 2.273 Gaps: 3
Percent Similarity: 66.087 Percent Identity: 31.304
1c_block:
56-991-2 x A1542894 ..
1/1 to: A1542894 from: 1 to: 696
rgAspGlnAspGly1LeTyrGlyLysAlaVal1LeI euAsnCys5c 712

```

```

111 .....
1 CCCACCGATAGAGCCTTCGCCAGGATCCGATGCCAAGGTGATGCA 50
712 rAlaGluGlyTyrProValProThr1LeVal1TrpLysPheSerLysGly 729
51 GCGTGATGGCTTCCCAACCCCAAGTTACTAGGAAGAACGCTGGCG 100
729 laGlyValProGlnPheGlnPro1LeAlaLeuAsnG1Arg1LeGlnVal 745
101 ACACCCCCGAGAGTACAAAGATCTAAGACAGGACGACAACTCCGCTG 150
746 LeuSerAsnGlySerLeuLeuLeuLysHisValValGluLysSerG1 762
151 ...GAAGAGGGCAGCTCGCATGCGACAAACATCCAAAGACCAAGAGG 197
762 yTyrTyrLeuGlyLysValSerAsnAspValAlaValAspValSerLys 779
198 CTACTATCTGTGGAGGCTATCATGATAGCTCCGCCCTCGGCTG 247
779 ermetyLLeuThrValLys1LeProAlaMet1LeThrSerTyrProAsn 795
248 TGATTATGATCAGCTTCAGGCCCTCCAGAGTTCAAGGAGAGCTGCGC 297
796 ThrThrLeuAlaThrGlnGlyGlnLysLysGluMetSerCysThrAla1 812
298 AACCAAGCCGCCGACGAGAGAACCCGCGTACTCCAGTCCGAGCGCAA 347
812 sGlyGlyLysPro1Le1LeValAlaGTrpGlyLysGluAspArg1Le1Le 829
348 GGGCGGAAGCCCATTTGGCATCTTGTGCAACATGAACAACATGCGACAC 397
829 snProLysMet...AlaArgTyrLeuValSerThrLysGluValGlyL 844
398 ACCCAAGAACGACATCGGTACACCATTCGTGAGAGATCTTTCACAC 447
845 GluVal1LeSerThrLeuG 1LeuProThrValArgGluAs SerG1 861
448 GCAGTATATGTCAGTCTGTCATCAAGCCGACGAGATCGCATCCGCG 497
861 yPhePheSerCysHisAla1LeAsnSerTyrGlyLysAspArgGly1LeI 878
498 CCTATTACCTGTGTGCGCACATATGCTTGTGATCCGATGCGACCA 547
876 leGlnLeuThrValGlnGluProProAspProPro...Glu1LeGlu1Le 893
548 TAAATATGATTGTCAGAGACTTTCGCGAUCATATGCTTTGAGGTA 597
894 LysAspValLysAlaArgThr1LeThrLeuArgTyrThrMetGlyPheAs 910
598 CTCGACAAATCCGAGAGCTTCGTCGAGCTGAGCTGGGGCAACCTTACGA 647
910 pGlyAsnSerPro1LeThrGlyTyrAsp1LeGluCysLys 923
648 TGGCAACTCCCTCTGGACAGGTACATCATGTAGTTAAG 687
seq_name: gb_est37:AW014717
seq_documentation_block:
LOCUS AW014717 409 bp mRNA EST 10-SEP-1999
DEFINITION UT-H-B10-aae-a-07-0-UT-s1 NCI-CGAP-SubI Homo sapiens cDNA clone
IMAGE:2709012 3', mRNA sequence.
ACCESSION AW014717
VERSION AW014717.1 GI:5863474
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```



COMMENT

On May 18, 1998 this sequence version replaced gi:1138023.  
 Contact: Robert Strausberg, Ph.D.  
 Tel.: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/ILM at:  
 www.bio.illn.gov/db/rlp/image/image.html The following repetitive  
 elements were found in this cDNA sequence: 1-21,  
 >AT richlow complexity  
 Seq primer: M13 Forward  
 POLY-A=yes.

FEATURES

Source  
 Location/Qualifiers  
 1..409  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2709012"  
 /clone\_1b="NCI CGAP Sub1"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Ph rmacia) with a modified  
 polylinker. Site 1: Not 1; Site 2: Eco RI; The  
 NCI CGAP Sub1 library is a subtracted library derived from  
 BI. BI constitutes a mixture of 21 normalized or  
 subtracted NCI CGAP libraries: NCI CGAP C04,  
 NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP C010,  
 NCI CGAP C016, NCI CGAP K1d5, NCI CGAP K1d12,  
 NCI CGAP K1d3, NCI CGAP K1d11, NCI CGAP Lym2,  
 NCI CGAP Br23, NCI CGAP C08, NCI CGAP CUL1, NCI CGAP Lel2,  
 NCI CGAP Br23, NCI CGAP Lu5, NCI CGAP Lu24,  
 NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6, NCI CGAP Br25.  
 These 21 libraries were pooled and a single-stranded DNA  
 preparation of the resulting mixture was used as a tracer  
 in a subtractive hybridization with a driver whose  
 composition is detailed below: NCI CGAP K1d3 pool 1 LAM  
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids  
 132376-132391, 1456008-1456775, 1500552-1502855)  
 NCI CGAP K1d5 pool 1 LAM 3338-3342, 3722-3725, 3776-3778  
 (IMAGE Clonoids 1323912-1325831, 1471368-1472903,  
 1492104-1493253) NCI CGAP Lu5 pool 1 LAM 3575-3582,  
 3851-3854 (IMAGE Clonoids 1414920-1417991,  
 1520904-1522439) NCI CGAP GC4 pool 1 LAM 3164-3167,  
 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,  
 1469064-1470983, 1475592-1476743) NCI CGAP Pr22 pool 1  
 LAM 2457-2459, 2758-2759, 3062, 368 (IMAGE Clonoids  
 985608-986759, 1101192-1101959, 1217928-1220615)  
 NCI CGAP C010 pool 1 LAM 2644-2653, 2871-2872 (IMAGE  
 Clonoids 1057416-1061255, 1144584-1145351) The resulting  
 subtracted library contained 530,000 recombinants.  
 Subtraction was performed as previously described  
 [Fonaldi, Lennon & Soares (1996): Normalization and  
 Subtraction: Two Approaches To Facilitate Gene Discovery.  
 Genome Research 6, 791-806.  
 TAG LIB=NCI CGAP K1d3  
 TAG TISSUE=Kidney  
 TAG\_SEQ=ATGC"  
 BASE COUNT 72 a 110 c 121 g 106 t  
 ORIGIN

alignment\_scores:

Quality: 344.00 Length: 100  
 Ratio: 3.954 Gaps: 0  
 Percent Similarity: 87.000 Percent Identity: 60.000

alignment\_block:

US-08-956-991-2 x AW014717/rev

Align seg 1/1 to reverse of: AW014717 from: 1 to: 409

404 AspGlyThrProLysIleLeSerAlaPheSerGluLysValSerPr 420

```

|||||
331 GATGGACGCCCGCATCGTCGTCCTTCAACGAGAAAGTGTCAACCC 282
420 oAlaGuproValSerLeuMeCysAsnValLysGlyThrProLeuProt 437
281 CGGGAGCGAGTTCTCTACTATGTGTGCGCCCAAGGCGCCCGCCCA 232
437 htlleThrPrlleuAspAspProLleuLysGlySerHis 453
231 CGGTCACTGGGCGCTCGACGATGACCCATGTCGGGATGCGACCC 182
454 ArgIleSerGlnMetIleThrSerGluLysAsnValSerTyrLeuAs 470
181 CGCACCAACAGTACACAT""CGAGCGCACACACATCAGCCACATGMA 132
470 nIleSerSerSerGlnValArgAspGlyValTyrArgCysThrAla 487
131 CGTCACAGCCCGCCAGATCCGCGCGGGCGGTGTACCGGTCCACAGCC 82
487 snAsnSerAlaGlyValLeuTyrGlnAlaArgIleAsnValArgGly 503
81 GGAACCTGTGTGGGACTGCTGATATACGCGCGATATAAGTAGAGGT 32
    
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:59:04 ; Search time 4987.54 seconds  
(without alignments)  
4854.765 Million cell updates/sec

Title: US-08-956-991-10  
Perfect score: 6413  
Sequence: 1 tgaactgagcgagcgcgcg.....gaaattgccaaaataatt 6413

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	qb_est1:*
21:	qb_est2:*
22:	qb_est3:*
23:	qb_est4:*
24:	qb_est5:*
25:	qb_est6:*
26:	qb_est7:*
27:	qb_est8:*
28:	qb_est9:*
29:	qb_est10:*
30:	qb_est11:*
31:	qb_est12:*
32:	qb_est13:*
33:	qb_est14:*
34:	qb_est15:*
35:	qb_est16:*
36:	qb_est17:*
37:	qb_est18:*
38:	qb_est19:*
39:	qb_est20:*
40:	qb_est21:*
41:	qb_est22:*
42:	qb_est23:*
43:	qb_est24:*
44:	qb_est25:*

45:	qb_est26:*
46:	qb_est27:*
47:	qb_est28:*
48:	qb_est29:*
49:	qb_est30:*
50:	qb_est31:*
51:	qb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	em_est33:*
60:	qb_est34:*
61:	qb_est35:*
62:	qb_est36:*
63:	qb_est37:*
64:	qb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	qb_est39:*
70:	qb_est40:*
71:	qb_est41:*
72:	qb_est42:*
73:	qb_est43:*
74:	qb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	qb_gss1:*
80:	qb_gss2:*
81:	qb_gss3:*
82:	qb_gss4:*
83:	em_gss1:*
84:	em_gss2:*
85:	em_gss3:*
86:	em_gss4:*
87:	qb_gss5:*
88:	qb_gss6:*
89:	qb_gss7:*
90:	qb_gss8:*
91:	qb_gss9:*
92:	em_gss5:*
93:	em_gss6:*
94:	em_gss7:*
95:	em_gss8:*
96:	em_gss9:*
97:	em_gss10:*
98:	em_gss11:*
99:	qb_gss10:*
100:	qb_gss11:*
101:	em_gss12:*
102:	qb_gss12:*
103:	qb_gss13:*
104:	qb_gss14:*
105:	qb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	415.8	6.5	439	43	AT243628	AT243628 qb88a03.x
C 2	399.2	6.2	541	64	AW046213	AW046213 UT-M-BH1-

C	3	397.4	6.2	419	42	A1094516	A1094516 ov63f11.s
C	4	374.2	5.8	510	25	N80326	N80326 yz91f03.r1
C	5	358.8	5.6	430	25	N64532	N64532 yz91f03.s1
C	6	343.8	5.4	355	37	AJ003472	AJ003472 AJ003472
C	7	341.2	5.3	458	46	A1454704	A1454704 UI-R-BT0
C	8	318	5.0	321	20	241519	241519 HSC2W042 n
C	9	303.4	4.7	307	41	A1051251	A1051251 oy49604.x
C	10	271.2	4.2	427	71	AM159035	AM159035 z850605.x
C	11	265	4.1	310	21	FL1426	FL1426 HSC2XA021 n
C	12	262.4	4.1	365	64	AM048531	AM048531 UI-M-BH1
C	13	260.4	4.1	294	20	245894	245894 HSC2W041 n
C	14	254.8	4.0	337	29	AA170935	AA170935 ms4505.x
C	15	242.4	3.8	252	21	F09100	F09100 HSC2XA022 n
C	16	213.4	3.3	264	46	A1448860	A1448860 ms4505.x
C	17	187	2.9	187	39	AA878733	AA878733 oe80601.s
C	18	175.6	2.7	619	79	FR0002941	286724 F.rubripes
C	19	163.2	2.5	402	64	AM048129	AM048129 UI-M-BH1
C	20	156	2.4	156	62	A1880033	A1880033 ap24b07.x
C	21	147.4	2.3	409	63	AM014717	AM014717 UI-H-B10
C	22	143	2.2	147	79	HSMC18B12	X88325 H.sapiens D
C	23	142.4	2.2	465	81	B07180	B07180 CSR172-h3-u
C	24	138.4	2.2	450	47	A1498283	A1498283 tm5h11.x
C	25	123	1.9	646	42	A1106884	A1106884 GH06134.5
C	26	122.8	1.9	602	79	FR0022038	AL014909 F.rubripes
C	27	120.6	1.9	442	79	FR0030733	AL027102 Fugu rubr
C	28	118.4	1.8	437	79	FR0002934	286717 F.rubripes
C	29	114.4	1.8	298	31	AA325603	AA325603 EST28915
C	30	111.4	1.7	315	104	A0552303	A0552303 RPCR-11-4
C	31	105.4	1.6	935	79	CNS006XK	AL066051 Drosoph11
C	32	104.6	1.6	282	71	AV264594	AV264594 AV264594
C	33	104	1.6	932	79	CNS00720	AL066742 Drosoph11
C	34	98.6	1.5	776	79	CNS010RX	AL093032 Drosoph11
C	35	97.4	1.5	935	79	CNS006XK	AL066051 Drosoph11
C	36	95	1.5	1101	80	CNS01523	AL041901 Drosoph11
C	37	94.8	1.5	1080	88	AO893056	AO893056 HS_4832.A
C	38	94	1.5	1101	80	CNS0153F	AL014949 Drosoph11
C	39	93.6	1.5	925	79	CNS0091P	AL053013 Drosoph11
C	40	93.6	1.5	1035	80	CNS017KO	AL108142 Drosoph11
C	41	92.8	1.4	99	79	HSMC34D09	X88135 H.sapiens D
C	42	92.6	1.4	925	79	CNS0091P	AL053013 Drosoph11
C	43	92.2	1.4	544	79	FR0022080	AL014951 F.rubripes
C	44	91.8	1.4	932	79	CNS00720	AL066742 Drosoph11
C	45	91.4	1.4	894	88	AO893147	AO893147 HS_3131.B

## ALIGNMENTS

RESULT 1  
LOCUS A1243628 439 bp mRNA EST 01-DEC-1998  
DEFINITION qb88a03.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens CDNA clone  
IMAGE:1854028 3', mRNA sequence.  
ACCESSION A1243628  
VERSION A1243628.1 GI:3839025  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sap. ens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 439)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 17, 1998 this sequence version replaced gi:2043430.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 476 Std Error: 0.00  
Seq primer: -40up from Gibco.

FEATURES  
source Location/Qualifiers  
1..439  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1854028"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="organ: pooled. Vector: pT73d-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NDH19W, testis NHT, and B-cell  
NCI CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
1.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 86 a 104 c 115 g 133 t 1 others  
ORIGIN  
Query Match 6.5%; Score 415.8; DB 43; Length 439;  
Best Local Similarity 98.1%; Pred. No. 1.2e-68;  
Matches 420; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 5986 ttctgttaacgaggtgtgtccagcaccagcagagacgtgacttaagacaacatgct 6045  
|||||  
DB 432 ttttttttTAACGAGGTGTCACGACACCCANCGACCTGAGTACGACATGCT 373  
QY 6046 tggaaacc cagaaaaacggagacctgaagcggccagctcttgagccatcccatg 6105  
|||||  
DB 372 TGGAAACC :AAGAAAACCCGACCTTGAAGCCGCCACGCTCTGAGCCATCCGATGG 313  
QY 6106 aagcgcctctctccgctctccacagagaagaacagtcgtgagcgggggctgctg 6165  
|||||  
DB 312 AAGCCCTCTCTCCGCTCTCTCCACGAGAGAGACAGCGTGGCAGCCGGGGCGCTGG 253  
QY 6166 cccattacttaagcgggggagcagagcgtgagcagcagcctaataatgacgctcc 6225  
|||||  
DB 252 CCACATTACTTACGCGGAGGAGGAGACCTTGGACAGCAGCTTAAATGACACTTCC 193  
QY 6226 aagaatcactgctgcagctccgcagccattgaagaagaacaatccttaacaaatct 6285  
|||||  
DB 192 AAGAACTACTGCTCGACTCCCGGGCCATTGAAAGGAACATCTTACCGAAATCTT 133  
QY 6286 acacccctgtataacagacagacatgactgagacgcggtgttaataatcaatcaaat 6345  
|||||  
DB 132 ACACCTGTATACAGACAGCATGACGACGCGTTGTAATATCAATCAACAAT 73  
QY 6346 caatcaaacctaccttttlltaaggaattccaattataatataaagaatgacaa 6405  
|||||  
DB 72 CAACTCAACCTACCTTTTTCGGAATTCGAATTTATTAATTAAGAAATTTGCCAA 13  
QY 6406 aatataat 6413  
|||||  
DB 12 AATATATT 5  
RESULT 2  
LOCUS AW046213 541 bp mRNA EST 18-SEP-1999  
DEFINITION UI-M-BH1-8kw-g-12-0-UI s1 NIH BMAP M.S2 Mus musculus CDNA clone  
ACCESSION AW046213  
VERSION AW046213.1 GI:5906742  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 541)

**AUTHORS** Bonaldo, M.F., Lennon, G. and Soares, M.B.  
**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery  
**JOURNAL** Genome Res. 6 (9), 791-806 (1996)  
**MEDLINE** 97044477  
**COMMENT** On Jun 5, 1998 this sequence version replaced g1:3188463.  
 Contact: Chn, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestrail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag served to identify it as a clone from the normalised hypothalamus library cDNA library preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence:  
 1-52, >At\_1rch#Low.complexity  
 Seq primer: M13 Forward  
 POLY-A-yes.

**FEATURES**  
**SOURCE** Location/Qualifiers  
 1..541  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U1-M-BH1-akw-g-12-0-U1"  
 /clone\_1lb="NIH\_BMAP\_M\_S2"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH\_BMAP\_M\_S2 library is a subtracted library derived from NIH\_BMAP\_M\_S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH\_BMAP\_M\_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.  
 TAG\_L1b=NIH\_BMAP\_M\_S2  
 TAG\_TISSUE=hypothalamus  
 TAG\_SEQ=CAGTA"

**BASE COUNT** 105 a 127 c 144 g 165 t  
**ORIGIN**

Query Match 6.2% Score 399.2; DB 64; Length 541;  
 Best Local Similarity 88.9% Pred. No. 1.7e-65;  
 Matches 466; Conservative 0; Mismatches 53; Indels 5; Gaps 3;

OY 5895 tcagatggaagaagatgaatgcaatgcaaaagga-atcgagccagtgagct 5953  
 ||||||| ||||| ||||| ||||||| ||||||| ||||||| |||||||  
 Db 541 TCAGGATGGAGGAGTGTGTAACATGGGGTTCCAAAGGCCCATCGCGACACT 482  
 OY 5954 catatctggcccatctactagatgagctttgttaaacagagtggtccagcac 6013  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 481 CATACACCTCTCCATACCTAGCAATGACTCTTGTTAAACGGGGGCGACACGAC 422  
 OY 6014 cagcagagacctgagcttgaacaacatgcttgaacctcaaaaagccgagacctga 6073  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 421 CAGGAGGACCTGAGTTTGAACAAGCGTCTTGGAACCCAGAAAAAGTCGACCTTGAA 362  
 OY 6074 ggcgcccaagctctgagaccctccgagtggaagccgctcctccgacctccacgag 6133  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 361 AGGCCACAGGCTCTGAGCCACCCCTATGAGAGGCCCTCTCTCCACCTTTCCACGC 302  
 OY 6134 agaaga---cagtcgagcgccgagggccgtggtgccaattaccagcgaggaagc 6190  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 301 AGAAGACAGCAGTCTGTGCAACAAGGGCTGTGGCCACTTACTCAGCAGAGGGTGC 242  
 OY 6191 aagagctggagcagcagcttaaatgagcagctcccaagaatcgtctgactccgagg 6250  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 241 AGAGCTGGACAGCAGCTTAATAATGACGACTCCCAAGAGTACTGCTGACCTCCGGG 182  
 OY 6251 ccattgaaaggaacaactcctcctacgcaaatcttaccacctgtgtataacgagaaga-t 6309  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 181 CCATTGGAAGGAACAATCCCTCCGCAAAATCTTACCTGGATTATACACATGCGACC 122  
 OY 6310 gactgagcagcggtgtgtaataatcaatcaatcaatcaatcaatcaatcaatcaatcaat 6369  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 121 TGATGACAGCGGTGTGTAATAATTTAAACGAGCAGCAATCAAGCTACTTTTATAC 62  
 OY 6370 ggaatccaatattatcaatlaaagaagaattgcgaatatatt 6413  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 61 TGATTCGATATTATATTAAAGAAATGCCCAGAAATATATT 18

**RESULT** 3  
 A1094516/c 419 bp mRNA EST 24-SEP-1998  
 LOCUS OY63f11.s1 NCI\_CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:1670541 3'  
**DEFINITION** mRNA sequence.  
 A1094516  
**ACCESSION** A1094516.1 GI:3433492  
**VERSION** EST.  
**KEYWORDS** SOURCE  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 419)  
 NCI/NIHNS-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index  
 Unpublished (1998)  
 On Jan 19, 1998 this sequence version replaced g1:2284845.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbp/image/image.html

**JOURNAL**  
**COMMENT**

**FEATURES**  
**SOURCE**

Insert length: 659 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham.  
 Location/Qualifiers  
 1..419  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1670541"  
 /clone\_1lb="NCI\_CGAP\_Brn23"  
 /tissue\_type="gliblastoma (pooled)"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer (5' TGTTACCAATCTGAAGTGGAGCGCGCATATCTTTTGTTTTGTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library is normalized, and was constructed by Bento







laetmid BA vector. Clone library from B. Soares, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S. in press"

BASE COUNT 66 a 70 c 83 g 99 t 3 others

ORIGIN

Query Match 5.0%: Score 318; DB 20; Length 321;

Best Local Similarity 99.1%: Pred. No. 3e-50;

Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

6093 cccatcccatggaagccgcctcctcgcctcctccacagagaagacagctgagcag 6152

321 cccatcccatggaagccgcctcctcgcctcctccacagagaagacagctgagcag 262

6153 cccgagggccgagccacatctcctcagcaggaagagagcctggagcagcctaaa 6212

261 cccgagggccgagccacatctcctcagcaggaagagagcctggagcagcctaaa 202

6213 atgagcagctcccaagaaatcactgctcagctcccgaggcatttgaagaagacaatcct 627

201 atgagcagctcccaagaaatcactgctcagctcccgaggcatttgaagaagacaatcct 142

6273 taacgcaaaatctacacccctggtatataacagagatgacatgagcagcgtgtgtaaatatc 6332

141 taacgcaaaatctacacccctggtatataacagagatgacatgagcagcgtgtgtaaatatc 82

6333 aattcaacaattcaatcaagcactcttcttctacggaattccaatattataatataa 6392

81 aattcaacaattcaatcaagcactcttcttctacggaattccaatattataatataa 22

6393 agaaaaattgcacaaatatact 6413

21 agaaaaattgcacaaatatact 1

RESULT 5

LOCUS A1051251 307 bp mRNA EST 24-SEP-1998

DEFINITION oy49e04.x1 NCI\_CGAP\_Brn23 Homo sapiens CDNA clone IMAGE:1669182 3'

ACCESSION A1051251

VERSION A1051251.1 GI:3306785

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 307)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BRGAP), Tumor Gene Index

Unpublished (1998)

On Jan 17, 1998 this sequence version replaced gi:2044499.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: koberl@strausberg.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Insert Length: 649 Std Error: 0.00

Seq primer: -40m3 fwd. ET from Amersham.

High quality sequence stop: 250.

Location/Qualifiers

1. .307

FEATURES

source

/organism="Homo sapiens"

/db\_xref="taxon:3606"

/clone="IMAGE:1669182"

/clone\_1b="NCI\_CGAP\_Brn23"

/tissue\_type="glioblastoma (pooled)"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pTZ19-Pac (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

tggtaccatctggaagtgagcggccgacatctgttttttttttttttttttttt

t 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTZ19 vector.

Library is normalized, and was constructed by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT 64 a 66 c 78 g 99 t

ORIGIN

Query Match 4.7%: Score 303.4; DB 41; Length 307;

Best Local Similarity 99.7%: Pred. No. 1.7e-47;

Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

6108 gccgcctcctccgctccctcccaagagaagacagctcgtggcagccggggcggtggcc 6167

307 gccgcctcctccgctccctcccaagagaagacagctcgtggcagccggggcggtggcc 248

6168 acattacctcagcgggaggaagcagcgtggagcagcagcctaaatgagcagctcccaa 6227

247 acattacctcagcgggaggaagcagcgtggagcagcagcctaaatgagcagctcccaa 188

6228 gaatcaatgctgagctcccgaggagcatttgaagaagaacatctcctacgcaaatcttacc 6287

187 gaatcaatgctgagctcccgaggagcatttgaagaagaacatctcctacgcaaatcttacc 128

6288 acccttgataacagacagcagcagcgtggagcagcgtgtgtaaatcaaatcaaatcaatca 6347

127 acccttgataacagacagcagcagcgtggagcagcgtgtgtaaatcaaatcaaatcaatca 68

6348 atcaaacctccttcttctacggaattccaatattataatgaagaagaatgccaaaa 6407

67 atcaaacctccttcttctacggaattccaatattataatgaagaagaatgccaaaa 8

6408 tatat 6412

7 TATAT 3

RESULT 10

LOCUS AM159035 427 bp mRNA EST 05-NOV-1999

DEFINITION za50e05.x1 Xenopus laevis EST library Xenopus laevis CDNA clone za50e05

5', mRNA sequence.

ACCESSION AM159035

VERSION AM159035.1 GI:6271064

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;

Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodidae;

Xenopus.

1 (bases 1 to 427)

Schultz,K., de la Bastide,M., Huang,E.N., Nascimento,L., Preston,R.,

Shah,R., Swaby,I., Shekher,M., Spiegel,L., Vill.M.D. and

McCombie,M.R.

Expressed sequence tags from Xenopus

Unpublished (1999)

On May 16, 1998 this sequence version replaced gi:3137856.

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

JOURNAL

COMMENT



Fax: 516 367 8874  
Email: mcmombe@cshl.org  
Plate: za50 row: e column: 05  
Seq primer: M13 universal forward primer  
High quality sequence stop: 427.  
Location/Qualifiers

FEATURES

source  
1. 427  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="za50e05"  
/clone\_lib="Xenopus EST library"  
/tissue\_type="total brain tissue"  
/cell\_line="W2-TGA"  
/dev\_stage="laddpole"  
/note="Vector: Lambda Zap I; Site-1: XbaI; This library was supplied by Holly Cline (Cold Spring Harbor Labs).  
cDNA synthesis with oligo df Xba I (Xba I cloning site).  
RNA: stage 50-56 tadpoles, total brain tissue, CTC extraction method."  
BASE COUNT 114 a 122 c 91 g 100 t  
ORIGIN

Query Match 4.2% Score 271.2; DB 71; Length 427;  
Best Local Similarity 79.8% Pred. No. 2e-41; Mismatches 78; Indels 7; Gaps 1;  
Matches 335; Conservative 0;

OY 1678 ttaattcctaccccttaagtaaaagtgtagtcacagacagcggttcctt..cgtgca 1737  
15 TCATCTCTGCATTAGCGAAGAGTGAGAGCCCGAGAACCTGTCTCTATGTGCA 74  
OY 1738 acgtgaaggaacaccccttgcgcacgacgtgagcccttgagatgaccgattcca 1797  
75 ACGTAAAGGAAGCGCTCTTCGAGATCCTTGAGACCTGGACGACGACCCCATATCA 134  
OY 1798 aggggtagcagtcacgcagcatcagatcagtcagtcgagaggaagtcgtagctacc 1857  
135 AGACGGCCACACAGATTAAGTCACTGATCCTTGAGGCAACGTTGACATAC 194  
OY 1858 tgaacatctcagctccagctccagtcgagagcgagtcacgcgcacgtccaaacac 1917  
195 TGAACATCACCAACTCAAGTCGAGATGCGGGGTGTACGATGACGCGTAAACACT 254  
OY 1918 cggcagagatctctcgttcacagctcgaataacgtaagagggcctcaagcattcgc 1977  
255 CTGGGGCGCTGCTTGTTCAGAGCTCGAATAACCT-----CCTGAATATTCGCG 307  
OY 1978 caatgaaacacatcacagcaatagcagagcgagacacatcattcactgctgctgattg 2037  
308 CAAATGAAAACATCACGCGCATAGCTGGCGGACACATACATCCACTGCCGCTCATTTG 367  
OY 2038 gctatcgtattctcatttaaatgtagtaagaactctaactgctccttccaaccac 2097  
368 GATATCCCTACTACTCATCAAGTGTACAAAACCTTAACCTCCCTCCATTTAATCACC 427  
Db

RESULT 11  
F13426 310 bp mRNA EST 15-MAR-1995  
LOCUS HSC2A021 normalized infant brain cDNA Homo sapiens cDNA clone  
DEFINITION C-2xa02, mRNA sequence.  
ACCESSION F13426  
VERSION F13426.1 GI:710043  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 310)  
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,  
Devignes, M. D., Duprat, S., Houllgate, R., Jumeau, M. N., Lamy, B.,  
Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y.,  
Sebastiani, Kabackchis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome  
and its expression  
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
MEDLINE 95277534  
COMMENT Contact: Genethon  
Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel.: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read.  
Genexpress library\_id: C; Genexpress\_sequence\_id: y2c-2xa02  
Insert length: 1500 Std Error: 0.00  
Seq primer: (-21)M13 universal  
High quality sequence stop: 335.  
Location/Qualifiers

FEATURES

source  
1. 310  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="C-2xa02"  
/clone\_lib="normalized infant brain cDNA"  
/sex="female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site-1: HindIII;  
Site-2: NotI; sex:Female; dev\_stage=3 months old;  
isolate-muscular atrophy patient; tissue\_type=total  
brain; total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmid BA vector. Clone library from B.Souares, Psychiatry  
Dept. Columbia University, USA. Normalization method:  
Bento Soares, P.N.A.S. In press"  
BASE COUNT 79 a 73 c 89 g 56 t 3 others  
ORIGIN

Query Match 4.1% Score 265; DB 21; Length 310;  
Best Local Similarity 95.4% Pred. No. 2.8e-40; Mismatches 13; Indels 0; Gaps 0;  
Matches 271; Conservative 0;

OY 5115 ggctcgcgaggaagacagcttaagagagctgcagatgcaagagtttcgtgaaatgct 5174  
27 GGAAGCGCGGAGAGAGGCTTAAGAGGCTGCACATGCAAGAGCTGUAATNCT 86  
OY 5175 catgagtaagaataccgacacilcagatacgtlaagcaagcaacagacacccctggaat 5234  
87 CATGACTGACATACCCGACCTT CAGTACGTTAACCAAGCAACAGACACCTCGCAAT 146  
OY 5235 gcaatcgaatccccagagctcagcttctgattgaagagagagacagatgagacat 5294  
147 GCACATGGCATACCCAGGCTCAGCTTTGATTGAAAGAGAGACACCATGAGACAT 206  
OY 5295 tgaatcagctccagagcttctgttgaagagatgctgagcatttgagagagcagtaagcaga 5354  
207 TGATGATGCTCCAGCGGTCTTNTTGACGAGATGCTGACTTTGGAGGGGCGACCTAACGAA 266  
OY 5355 gtccctgagcgtcactcaacagcgttcacttaacaatcggtcttc 5398  
Db 267 GTCCCTGACGGTCACTCACAGCGGTCCATTACCAATCGGTCTTC 310

RESULT 12  
AM048531/c 365 bp mRNA EST 18-SEP-1999  
LOCUS UI-M-BH1-ax-b-07-0-UI.s1 NIH-BMAP\_M\_S2 Mus musculus cDNA clone  
DEFINITION UI-M-BH1-ax-b-07-0-UI 3', mRNA sequence.  
ACCESSION AM048531  
VERSION AM048531.1 GI:5909060  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;



Qy 5170 atgctatgataagaatacccgacttcagatgacgaagaacagacacctg 5229  
|||||  
Db 73 ATGCTATAGTAAGANTACCCGAGCTTACGTTAGACAGACAGACCTG 132  
Qy 5230 cgaatgcacatgacataccagggctcagcttctgattgaagaagaacagctgag 5289  
|||||  
Db 133 CCAATGCACATGCNATACCAGGGCTCAGCTTTGATTGAAGAGACAGACGATGAG 192  
Qy 5290 accattgaatgactcgtccaggtctctgtgaagagatgacttggagaggaactaag 5349  
|||||  
Db 193 ACCATTGATGATCGCTCCAGGCTCTNTTGAAGATGCTGACTTTGGAGAGCAGCTAAG 252  
Qy 5350 cagaagctccctacagctcactcacacggtccattaccatcg 5391  
|||||  
Db 253 CAGAAGTCCCTGACGCTCACTCACAGGCTCATTTACCATCG 294

## RESULT 14

AA170935 337 bp mRNA EST 23-DEC-1996  
LOCUS ms45h05.r1 life Tech mouse embryo 13 5dpc 10666014 Mus musculus  
DEFINITION CDNA clone IMAGE:614553 5', mRNA sequence.  
ACCESSION AA170935  
VERSION AA170935.1 GI:1752261  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 337)  
AUTHORS Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The Washu-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 9, 1995 this sequence version replaced g1:802244.  
Contact: Matra M/Mouse EST Project

Washu-HHMI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LIND : contact the  
IMAGE Consortium (info@image.lind.gov) for further information.  
MGI:375377

Putative full length read  
vector to vector length is 479  
Seq primer: -28M13 rev1 from Amersham  
High quality sequence stop: 325.

## FEATURES

Location/Qualifiers

1..337

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="614553"  
/clone\_lib="Life Tech mouse embryo 13 5dpc 10666014"  
/tissue\_type="embryo"  
/dev\_stage="13.5dpc embryos"  
/lab\_host="DH10B"  
/note="Organ: whole embryo; Vector: pCMV-SPORT2; Site:1;  
Salt: Site:2; Notti: Cloned unidirectionally. Primer:  
Oligo dr. 13.5dpc embryos. pCMV-SPORT2 vector."

BASE COUNT 85 a 108 c 90 g 54 t

## ORIGIN

Query Match 4.0%: Score 254.8; DB 29; Length 337;  
Best Local Similarity 86.9%: Pred. No. 2.4e-38;  
Matches 292; Conservative 0; Mismatches 42; Indels 2; Gaps 1;

Qy 5943 ccaagtgacctacatacttgcctccataccttagaatgacttttgttaaccgaggt 6002  
|||||  
Db 2 CCAAGGACACCTATACACCTGCTCCATACCTAGCAATGGACTCTTGTAAACCGGGGC 61  
Qy 6003 ggtccagagccacgacgaggaacctgaagcaagaacatgcttggaaacataaagc 6062  
|||||  
Db 62 GCACACGACACACGAGGAGGAGCTGATTAGGACAGCGCTGTTGGAAACCCAGAAAGT 121  
Qy 6063 cgaacctgaagcgccacacggtcctgagccatcccgatggaagcgccctccgcgc 6122  
|||||  
Db 122 CGGACCTGTGAAGAGCCGCCACAGCTTCATGAGCCACCCCTATGAGGCTCTCTCCAC 181  
Qy 6123 tctccacgagaggaaggaagtcgtgagcagc--cgaggcgctggccacatcactcagc 6180  
|||||  
Db 182 TCTTCCACGCGAAGAGAGACAGCAGTGTGGCAACAAGGGGCTGTGGACCTTACTCAGC 241  
Qy 6181 gggaggagcagagcgtggagacagcgctaaataagcagcgtcccaagaatcactgctg 6240  
|||||  
Db 242 GAGAGGCTGACAGCTGGAGCAGCGCTAAATGAGCAGCTCCAGAGCTCAGTGG 301  
Qy 6241 actccggggccacttgaagaagaacaatccttc-g 6276  
|||||  
Db 302 ACTCCGGGGCCATTGGAAGGAACAATCCCTACG 337

## RESULT 15

F09100/c 252 bp mRNA EST 23-FEB-1995  
LOCUS HSC2XA02 normalized infant brain cDNA Homo sapiens CDNA clone  
DEFINITION c-2xa02 3', mRNA sequence.  
ACCESSION F09100  
VERSION F09100.1 GI:678256  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS

Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,  
Devignes,M.D., Duprat,S., Houllgate,R., Jumeau,M.N., Lamy,B.,  
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouliot,Y.,  
Sebastien-Kabaktchis,C. and Tessier,A.  
IMAGE: molecular integration of the analysis of the human genome  
and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
95277534

## COMMENT

JOURNAL MEDLINE

Contact: Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex. FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. Removed at sequence 5'end  
Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: alc-2xa02  
Insert length: 676 Std Error: 0.00  
Seq primer: (-21)M13 universal  
High quality sequence stop: 268.

## FEATURES

Location/Qualifiers

1..252

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-2xa02"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: latmid BA; Site:1; HindIII:  
Site:2; Notti: sex:Female; dev-stage=3 months old;  
isolate=muscular atrophy patient; tissue-type=total  
brain; total mRNA was oligo (dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> Notti sites of the  
latmid BA vector. Clone library from B.Souares, Psychiatry

Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S. in press"  
BASE COUNT 58 a 50 c 53 g 83 t 8 others  
ORIGIN

Query Match 3.8%; Score 242.4; DB 21: Length 252;  
Best Local Similarity 96.4%; Pred. No. 4.9e-36;  
Matches 243: Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Oy 5162 ytgccacattacc:ccagccggaggagagagagctggagacagcctaataatgagcagc 6221  
|||||  
Db 252 GTGGCCACATTACTCTCAGCGGAGGAGCAGAGCTGGACAGCGCAGCTNAAATGAGCAGC 193  
6222 lcccaagaatcactgcctcgaactcccgaggccattgaaagaaacaatccttacgcaaa 6281  
|||||  
Db 192 TCCCNAGATCAGCTGCTGCACTCCCGGGCCATTGAAAGAAACAATCCTTACGGAAAA 133  
Oy 6282 tcttaacacctggtataaagacagacatgactggacagcggtgttaaatlacaattcaaac 6341  
|||||  
Db 132 TCTTACACCTGTGATMACAGACGACATGACTGACAGNGGTTGTAATPACAAATCAAAC 73  
Oy 6342 aattcaatcaaagctaaccttttttaacggaattccaatattataatataaagaatgtg 6401  
|||||  
Db 72 AANTCAATNNAAGCTACTTTTTTTACGGAATTCATATTTATATTAAGNAAATTG 13  
Oy 6402 ccaaatatatt 6413  
|||||  
Db 12 CCNNAATATATT 1

Search completed: April 25, 2000, 16:59:22  
Job time: 7998 sec



FT Disulfide\_bond 428..484  
FT Disulfide\_bond 525..575  
FT Disulfide\_bond 617..669  
FT Disulfide\_bond 711..766  
FT Disulfide\_bond 809..865  
FT Disulfide\_bond 1307..1359  
FT Modified\_site 78..80  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 106..108  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 470..472  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 487..489  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 658..660  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 666..668  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 710..712  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 748..750  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 795..797  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 924..926  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1142..1144  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1160..1162  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1250..1252  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1271..1273  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1324..1326  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1341..1343  
FT Modified\_site /note="Asn is N-glycosylated"  
FT Modified\_site 1488..1490  
FT Modified\_site /note="Asn is N-glycosylated"  
PN MO9817795-A1.  
PN 30-APR-1998.  
PR 23-OCT-1997: U19547.  
PR 25-OCT-1996: US-029322.  
PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
PI Korenberg JR:  
PI WP1: 98-271791/24.  
DR N-PSDB: V31981.  
PT New isolated Down's Syndrome-cell adhesion molecule - used to  
PT develop products for detection, diagnosis and therapy of  
PT developmental and neurological abnormalities  
PS Claim 2: Page 73-78: 109pp: English.  
CC This polypeptide comprises Down syndrome-cell adhesion molecule  
CC DS-CAM1, a cell surface glycoprotein belonging to a novel subclass  
CC of the Ig superfamily with highest homology to neural cell adhesion  
CC molecules. Its amino acid sequence was deduced from cDNA clones  
CC (see V31981) isolated from a trisomy 21 foetal brain library. A  
CC splice variant, DS-CAM2 (see M42087), which is non-membrane bound  
CC was also identified. The invention also provides human and murine  
CC DS-CAM nucleic acid sequences (see also V31981-88), expression  
CC vectors and host cells, transgenic animals, antibodies, antisense  
CC oligonucleotides, and primers derived from DS-CAM nucleic acid.  
CC DS-CAM polypeptides are associated with developmental and  
CC neurological processes. They can be used in e.g. neural prosthetic  
CC devices used in entubulation methods of repairing (regenerating)  
CC damaged or severed peripheral nerves, and also in bioassays to  
CC identify agonists and antagonists. The products can also be  
CC used in detection, diagnosis and therapy of developmental and  
CC neurological abnormalities such as Down syndrome, mental  
CC retardation, holoprosencephaly, agenesis of the corpus callosum,  
CC or schizencephaly.  
CC Sequence 1910 AA:

Query Match 100.0%; Score 13516; DB 1; Length 1910;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 1910; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 MMIALSLFQSFANFSEDLHSLLYFNASLQEVVFASTGTGLVPCAAACIPVTLRWYL 60  
1 MMIALSLFQSFANFSEDLHSLLYFNASLQEVVFASTGTGLVPCAAACIPVTLRWYL 60  
Db 61 ATGEETLYDPGIRHVNPGTLDLPFPSSSTLIHDTTYCTAENNSGKIRSDVHIKA 120  
61 ATGEETLYDPGIRHVNPGTLDLPFPSSSTLIHDTTYCTAENNSGKIRSDVHIKA 120  
Qy 121 VLREPTVREDDQKTRGNVAVFKCIIPSSVEAYITVVSWEKDTVLSGSRFLITSTGA 180  
121 VLREPTVREDDQKTRGNVAVFKCIIPSSVEAYITVVSWEKDTVLSGSRFLITSTGA 180  
Db 181 LYTKDVQNEDEGLYNYCITRRHRYTGETRQSNARLFVSDPNASPSILDGFDHKKAMAG 240  
181 LYTKDVQNEDEGLYNYCITRRHRYTGETRQSNARLFVSDPNASPSILDGFDHKKAMAG 240  
Qy 241 RVELPCALGHPEDYRMLKDNMPLELSGRFQKTYTGLLIENIRPSDGSYCEVSNRYG 300  
241 RVELPCALGHPEDYRMLKDNMPLELSGRFQKTYTGLLIENIRPSDGSYCEVSNRYG 300  
Db 301 TAKYIGRLYKQPLKATISPRKYSVGSVGLSCSYGTEDEQELSWYRNGEILNPKNV 360  
301 TAKYIGRLYKQPLKATISPRKYSVGSVGLSCSYGTEDEQELSWYRNGEILNPKNV 360  
Qy 361 RITGINHENLIMDMVNSDGAQVCFVRKDKLSAODYVOVLEDDPKIISAFSEKRVSP 420  
361 RITGINHENLIMDMVNSDGAQVCFVRKDKLSAODYVOVLEDDPKIISAFSEKRVSP 420  
Db 421 AEPVSLMCNVKGTPLTIYTLDDDLPLKGGSHRISOMITSEGNVSYNISSQVRODG 480  
421 AEPVSLMCNVKGTPLTIYTLDDDLPLKGGSHRISOMITSEGNVSYNISSQVRODG 480  
Qy 481 AEPVSLMCNVKGTPLTIYTLDDDLPLKGGSHRISOMITSEGNVSYNISSQVRODG 480  
481 AEPVSLMCNVKGTPLTIYTLDDDLPLKGGSHRISOMITSEGNVSYNISSQVRODG 480  
Db 481 VYRCANNSAGVYLYOARINVRGPASIRPKNITAIAGRDYIHCNAVIGPYYSIKWYN 540  
481 VYRCANNSAGVYLYOARINVRGPASIRPKNITAIAGRDYIHCNAVIGPYYSIKWYN 540  
Qy 541 SNLPLFNHROVAFFENGTGLKLSVDQKEVDEGTYCNLVQPOLSTOSVHYVTKVPEFI 600  
541 SNLPLFNHROVAFFENGTGLKLSVDQKEVDEGTYCNLVQPOLSTOSVHYVTKVPEFI 600  
Db 601 PREPFRFSGQVFIQCVVSGDLPITITWOKDGRPIPGSLGYTIDNIDFTSLRLSNLS 660  
601 PREPFRFSGQVFIQCVVSGDLPITITWOKDGRPIPGSLGYTIDNIDFTSLRLSNLS 660  
Qy 661 LMHNGNYTCIARNEAAVEHQSOLIVRPKFEVQPRDODIYKAVILNCSAEGYVPPT 720  
661 LMHNGNYTCIARNEAAVEHQSOLIVRPKFEVQPRDODIYKAVILNCSAEGYVPPT 720  
Db 721 IYWKFSKGGVYQFOPIALNGRIQVLSNGSLLIKHYVEDSGYLCKVSDVADYSKM 780  
721 IYWKFSKGGVYQFOPIALNGRIQVLSNGSLLIKHYVEDSGYLCKVSDVADYSKM 780  
Qy 781 YLTKVFKIPAMTSPNTTLATQCKEMSCIAHGKPIIVREKEDRIINPMARYLVSTK 840  
781 YLTKVFKIPAMTSPNTTLATQCKEMSCIAHGKPIIVREKEDRIINPMARYLVSTK 840  
Db 841 EVGEVISTQILIPYREDSGFSCHAINSYGEDGIIQLYVOEPPPEIEIDVART 900  
841 EVGEVISTQILIPYREDSGFSCHAINSYGEDGIIQLYVOEPPPEIEIDVART 900  
Qy 901 ITLRMTGFPNGSPIIGYDECKNKSQSDMSAQRTKDVSPQLNSATIIDIHPSSTYSIRM 960  
901 ITLRMTGFPNGSPIIGYDECKNKSQSDMSAQRTKDVSPQLNSATIIDIHPSSTYSIRM 960  
Db 961 YAKNRICKSEPSNELITADDEAAPDPPQVHLEPISSOSIVTKAKPKHLONGIINGY 1020  
961 YAKNRICKSEPSNELITADDEAAPDPPQVHLEPISSOSIVTKAKPKHLONGIINGY 1020  
Qy 1021 QIGYREVSTGNGFNFIISVDTSGDSEVYTLDMLNKFTQYGLVVOACNRAGTGPSSOETI 1080

\*\*\*\*\*  
M P E S E R E H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MSPrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Apr 25 13:02:51 2000; Maspar time 28.68 Seconds  
662.987 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-956-991-2  
Description: (1-1910) from US08956991A.pep  
Perfect Score: 13516  
Sequence: 1 MMTLALSLTQSFANVSEDL.....KAIGVTSYICHTLEWTF 1910

Scoring table:  
PAM 150  
Gap 11

Searched: 131253 seqs, 12956647 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 39.388; Variance 240.403; scale 0.164

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2081	15.4	465	2	US-08-752-	Sequence 5, Applicatio	2.74e-132
2	2071	15.3	462	2	US-08-752-	Sequence 7, Applicatio	1.37e-131
3	740	5.5	1447	3	PCT-US94-0	Sequence 2, Applicatio	2.64e-19
4	629	4.7	1018	1	US-08-408-	Sequence 6, Applicatio	8.67e-32
5	629	4.7	1018	1	US-08-408-	Sequence 6, Applicatio	8.67e-32
6	629	4.7	1018	1	US-08-714-	Sequence 6, Applicatio	8.67e-32
7	606	4.5	1018	1	US-08-452-	Sequence 2, Applicatio	3.07e-30
8	569	4.2	1911	1	US-08-348-	Sequence 5, Applicatio	9.38e-28
9	569	4.2	1911	2	US-08-800-	Sequence 5, Applicatio	9.38e-28
10	569	4.2	1911	3	PCT-US94-1	Sequence 5, Applicatio	9.38e-28
11	486	3.6	1501	2	US-08-716-	Sequence 3, Applicatio	3.23e-22
12	486	3.6	1501	2	US-08-447-	Sequence 3, Applicatio	3.23e-22
13	449	3.3	596	2	US-08-752-	Sequence 13, Applicati	9.06e-16
14	448	3.3	612	2	US-08-752-	Sequence 11, Applicati	1.05e-19
15	448	2.9	607	2	US-08-752-	Sequence 12, Applicati	5.78e-16
16	381	2.8	605	2	US-08-752-	Sequence 8, Applicatio	2.59e-15
17	379	2.8	946	3	PCT-US95-0	Sequence 13, Applicati	3.48e-15
18	361	2.7	615	2	US-08-752-	Sequence 9, Applicatio	5.13e-14
19	359	2.7	630	2	US-08-752-	Sequence 14, Applicati	6.92e-14
20	322	2.4	478	3	PCT-US95-0	Sequence 15, Applicati	1.66e-11
21	322	2.4	860	3	PCT-US95-0	Sequence 19, Applicati	1.66e-11
22	322	2.4	868	3	PCT-US95-0	Sequence 21, Applicati	1.66e-11
23	320	2.4	868	1	US-08-374-	Sequence 1, Applicatio	2.23e-11

24	320	2.4	868	2	US-08-644-	Sequence 1, Applicatio	2.23e-11
25	326	2.4	869	1	US-08-374-	Sequence 16, Applicati	9.20e-12
26	326	2.4	869	2	US-08-644-	Sequence 29, Applicati	9.20e-12
27	282	2.1	2231	1	US-08-153-	Sequence 16, Applicati	5.74e-09
28	262	1.9	2324	1	US-08-283-	Sequence 1, Applicatio	1.03e-07
29	262	1.9	2324	3	PCT-US95-0	Sequence 1, Applicatio	1.03e-07
30	262	1.9	2337	4	5455158-1	Patent No. 5455158.	1.03e-07
31	262	1.9	2386	2	US-09-016-	Sequence 12, Applicati	1.03e-07
32	262	1.9	2446	2	US-08-551-	Sequence 2, Applicatio	1.03e-07
33	262	1.9	2446	3	PCT-US93-1	Sequence 2, Applicatio	1.03e-07
34	233	1.7	252	2	US-08-414-	Sequence 57, Applicati	6.38e-06
35	233	1.7	287	2	US-08-414-	Sequence 49, Applicati	4.81e-06
36	229	1.7	287	2	US-08-414-	Sequence 48, Applicati	1.12e-05
37	229	1.7	304	2	US-08-414-	Sequence 44, Applicati	1.12e-05
38	229	1.7	310	2	US-08-414-	Sequence 45, Applicati	4.81e-06
39	235	1.7	315	2	US-08-414-	Sequence 47, Applicati	4.81e-06
40	235	1.7	338	2	US-08-414-	Sequence 43, Applicati	4.81e-06
41	235	1.7	338	2	US-08-414-	Sequence 42, Applicati	4.81e-06
42	234	1.7	419	4	5169835-2	Patent No. 5169835.	5.54e-06
43	232	1.7	549	2	US-08-836-	Sequence 11, Applicati	7.35e-06
44	232	1.7	574	2	US-08-836-	Sequence 21, Applicati	7.35e-06
45	234	1.7	828	1	US-08-261-	Sequence 2, Applicatio	5.54e-06

## ALIGNMENTS

RESULT 1  
ID US-08-752-307B-5 STANDARD: PRT: 465 AA.

AC xxxxxx

Sequence 5 Application US/08752307B  
Patent No. 5952171

GENERAL INFORMATION:

APPLICANT: McCarthy, Sean A.

APPLICANT: Geising, David P.

APPLICANT: Levinson, Douglas A.

TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/752.307B

FILING DATE: 19-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Melkijohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 09404/020001

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 465 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC FRAGMENT TYPE: Internal  
CC SEQUENCE 465 AA: 51210 MW: 1158127 CN:

Query Match 15.4% Score 2081; DB 2; Length 465;  
Best Local Similarity 57.1%; Pred. No. 2,74e-132;  
Matches 265; Conservative 105; Mismatches 93; Indels 1; Gaps 1;

Db 1 MMLVTFLLDLSLHKARPEVDGTSLYFVNDLSLOQVTFSSVGVVVPCCPAGSPSALRMW 60  
1 MWLLA-LSLFSOFANVFSEDLHSLFVNASLOEVFASTGTLPVCPAAGIPVTLRMW 59  
Cc 1 LATGDDIYDVPHIRVHANGTLQLYPFSPAFNSFIHNDYFCTAENAGKIRSPNIRVK 120  
Cc 50 LATGEIYDVPGIRVHANGTLQIFPPSSFTLIHNTYCTAENPSGKIRSDVHIK 119  
Cc 121 AVFREPTVAVEDQSRMGNVAVFKCLIPSSVOEYVSVSMKDTVS-IPENREFITYHG 180  
Cc 120 AVFREPTVAVEDQSRMGNVAVFKCLIPSSVEAYITVSMKDTVS-LVSGSRFLITSTG 179  
Cc 181 GLYISDVOKEDALSTYRCITKHKYSGETROSNGARLSTDPAESIPTILDGFHSGOEVMAG 240  
Cc 180 ALYIKDVONEDGLYNYRCTIRHRYTGETROSNSARLFVSDPANSAPSLDGFDRKKAMAG 239  
Cc 241 HTVELPCTASGYPPIAIRWLKDGRLPADSRMTKRTITGLTISDLRTEDSGITYICEVTNTE 300  
Cc 240 QVVELPCALGHPEPDYRMLKDNMPLSLGRFOKTYVTGLIENIRPSDSGSI-EVSNRY 299  
Cc 301 GSAEATGILMWIDPLVHTLTPPKLKTGISTVLSLCALTGSEPTIRMYRNTEVLVLPDEA 360  
Cc 300 GTRAKVIGRLVYKOPKATISPRKVSXSVGSVLSCTVGTEDQLSWYRNGEILNPGKN 359  
Cc 361 ISIRGLSNETLITSAQKSHSGAYOCFATRKQTAODFAIILAEDEGTIRIYSSSEKVVN 420  
Cc 360 VRTITGINHENLIMDHVKSXSGAYOCFVRKDKLSAODYVQVLEDTGPKIISAFSEKVV 419  
Cc 421 PGDFSLMCAAKGAPPPTYMALDDEPIYRDSGSHRTNYSMDG 464  
Cc 420 PAEPVSLMCNVKGPPLTIITWLLDDPILKGSGRHSISQMIISG 463

RESULT 2  
ID US-08-752-307B-7 STANDARD: PRT: 462 AA.  
XX xxxxxx  
AC  
XX  
DT  
XX  
DE Sequence 7, Application US/08752307B  
XX  
CC Sequence 7, Application US/08752307B  
CC Patent No. 5952171  
CC GENERAL INFORMATION:  
CC APPLICANT: McCarthy, Sean A.  
CC APPLICANT: Geary, David P.  
CC APPLICANT: Levinson, Douglas A.  
CC TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES  
CC TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Fish & Richardson, P.C.  
CC STREET: 225 Franklin Street  
CC CITY: Boston  
CC STATE: MA  
CC COUNTRY: US  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: Windows95

CC SOFTWARE: FastSeq for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/752.307B  
CC FILING DATE: 19-NOV-1996  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Melkiojohn, Ph.D., Anita L.  
CC REGISTRATION NUMBER: 35,283  
CC REFERENCE/DOCKET NUMBER: 09404/020001  
CC TELEPHONE: 617-542-5070  
CC TELEFAX: 617-542-8906  
Cc 200154  
Cc INFORMATION FOR SEQ ID NO: 7:  
Cc SEQUENCE CHARACTERISTICS:  
Cc LENGTH: 462 amino acids  
Cc TYPE: amino acid  
Cc TOPOLOGY: linear  
Cc MOLECULE TYPE: protein  
Cc SEQUENCE 462 AA: 50937 MW: 1144658 CN:

Query Match 15.3% Score 2071; DB 2; Length 462;  
Best Local Similarity 57.4%; Pred. No. 1.37e-131;  
Matches 263; Conservative 104; Mismatches 90; Indels 1; Gaps 1;

Db 1 MMLVTFLLDLSLHKARPEVDGTSLYFVNDLSLOQVTFSSVGVVVPCCPAGSPSALRMW 60  
1 MWLLA-LSLFSOFANVFSEDLHSLFVNASLOEVFASTGTLPVCPAAGIPVTLRMW 59  
Cc 1 LATGDDIYDVPHIRVHANGTLQLYPFSPAFNSFIHNDYFCTAENAGKIRSPNIRVK 120  
Cc 60 LATGEIYDVPGIRVHANGTLQIFPPSSFTLIHNTYCTAENPSGKIRSDVHIK 119  
Cc 121 AVFREPTVAVEDQSRMGNVAVFKCLIPSSVOEYVSVSMKDTVS-IPENREFITYHG 180  
Cc 120 AVFREPTVAVEDQSRMGNVAVFKCLIPSSVEAYITVSMKDTVS-LVSGSRFLITSTG 179  
Cc 181 GLYISDVOKEDALSTYRCITKHKYSGETROSNGARLSTDPAESIPTILDGFHSGOEVMAG 240  
Cc 180 ALYIKDVONEDGLYNYRCTIRHRYTGETROSNSARLFVSDPANSAPSLDGFDRKKAMAG 239  
Cc 241 HTVELPCTASGYPPIAIRWLKDGRLPADSRMTKRTITGLTISDLRTEDSGITYICEVTNTE 300  
Cc 240 QVVELPCALGHPEPDYRMLKDNMPLSLGRFOKTYVTGLIENIRPSDSGSI-EVSNRY 299  
Cc 301 GSAEATGILMWIDPLVHTLTPPKLKTGISTVLSLCALTGSEPTIRMYRNTEVLVLPDEA 360  
Cc 300 GTRAKVIGRLVYKOPKATISPRKVSXSVGSVLSCTVGTEDQLSWYRNGEILNPGKN 359  
Cc 361 ISIRGLSNETLITSAQKSHSGAYOCFATRKQTAODFAIILAEDEGTIRIYSSSEKVVN 420  
Cc 360 VRTITGINHENLIMDHVKSXSGAYOCFVRKDKLSAODYVQVLEDTGPKIISAFSEKVV 419  
Cc 421 PGDFSLMCAAKGAPPPTYMALDDEPIYRDSGSHRTNYSMDG 464  
Cc 420 PAEPVSLMCNVKGPPLTIITWLLDDPILKGSGRHSISQMIISG 457

RESULT 3  
ID PCT-US94-05277-2 STANDARD: PRT: 1447 AA.  
XX xxxxxx  
AC  
XX  
DT  
XX  
DE Sequence 2, Application PC/TUS9405277  
XX  
XX Sequence 2, Application PC/TUS9405277  
CC GENERAL INFORMATION:  
CC APPLICANT: Bruskin, Arthur





[illegible][illegible]



Query Match	99.9%	Score 6599.2;	DB 1;	Length 6604;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 6601:	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;

[illegible]

QY	1021	atgatattgtataactaccgctgtgataccggcatcgtatcaaccggagaagacgagcaga	1080
DB	1021	ATGATATTGTATACTACCGCTGTGATACCGGCATCTGATACCGGAGACGAGCGAGA	1080
QY	1081	gcaacagcgcgaagacttttgtatacagaccagcgaaactcgaaccccatcatalcayagt	1140
DB	1081	GCAACAGCGCGAAGACTTTTGTATCAGACCCAGCCGCACTCACCCCATTCATCTGATG	1140
QY	1141	ggtttgaccatcgcaaaagccatgagctgggcagcgigtgagactgtccttgcaagcctcg	1200
DB	1141	GGTTTGACCATTCGCAAAAGCCATGGCTGGGACAGGTGTGAGCTGCTTGGCAAGGGCTCG	1200
QY	1201	ggagacctgagccaagatctacgcctgtgataaagaazaactgtcccttggaactttcagga	1260
DB	1201	GGACCTTGAGCCAGATTACCGCTGTGCTAGAGACACATGCCCCGTGGAACTTTCAGGGA	1260
QY	1261	ggttcacgaagaccgttgacgaggctgtcatctgagaacatlcgcgccctcggactcaagca	1320
DB	1261	GGTTCACGAAGACCGTGACGGGCGCTGCTCATTTGAGAACATTCGGCCCTCGGACTCAGCA	1320
QY	1321	gctatgttttgaaagtgtgtccaaagaatacggaaactgctaaagtgtataagccgctgtatcg	1380
DB	1321	GCTATGTTTGTGAAGGTGTCCAAACGATACGGAACTGTAAAGTGTATAGGCCGCTGTACG	1380
QY	1381	tgaaacagccacgtgaaagcccaccatcactcaccaggaagaattaaagcagctgtgatacc	1440
DB	1381	TGAACAGCCCACTGAAAGCCACCATCATGTCCTCCAGAGAGTTAAAGCAGCGTGGGTACC	1440
QY	1441	aagtttccttgtctctgcgcgcgtgacagaactgtagagccaggaaactctcctgtatccga	1500
DB	1441	AAGTTTCTTGTCTCTCTGCGAGCTGCACAGGAACCTGAGACCGAGAACCTCTCTGTACCCCA	1500
QY	1501	atgtgtgaatccctcaaacctcggaaataatgtataagatacacaaggaatacaacacgaataac	1560
DB	1501	ATGTGTGAATCTCTCAACCTCGGAATAAATGTAGATCTCACAGGATCAACCAAGAAATCC	1560
QY	1561	ttataatgatatcacatgattcctaaagtacacggaggcgcatatacagttgtttgtgcgaag	1620
DB	1561	TTATAATGATATCATATGTTCAAAAGTAGCGGGGGCCATATACAGTGTCTTGTGCCCAAG	1620
QY	1621	acaagcgtgcgcgtccaagatatgtgcaggtgtgcttgaataatgaactcccataata	1680
DB	1621	ACAAAGCTGTCCGCTCAAGACTATGTGCAGGTGTCTTGAATATGAATCTCCAAATAATTA	1680
QY	1681	ttctctgctttagtgaaaaagttgttgagtcacagcagagccggtttccctatgtgcacag	1740
DB	1681	TTTCTGCTTTAGTGAATAAGTGCTGTAGTCCAGACAGCCGGTTTCCCTATGTGTACAG	1740
QY	1741	tgaaggagaacaccttgcgcacgaatacatttgaaaccttgagcgaatacccgatctcaag	1800
DB	1741	TGAAGGAGAACCTTTGCCCCACGATACGTGTGACCTGTGACATATACCCGATTTCTCAAG	1800
QY	1801	gttgacgtacacgcgatcatcagcagatgatcaagtcggaaggagagctgtgtcagatctcga	1860
DB	1801	GTTGACGTACACCGCATATAGCCCATATATCATCGTGGAGGGAACTGTGTCACTACTCTGA	1860
QY	1861	acatcccaagctcccaaggtcccggaacggaggagtctacccgtgcacgtgcacaaacactcgg	1920
DB	1861	ACATCTCCAGCTCCACAGTCTCGGACGGGGAGTCTACCGCTGTCACTGCCAAACACTCGG	1920
QY	1921	cgggaagtcgtctcttaccagagctcgaataaagctlaagagaggtcgtcaagcatctgaacca	1980
DB	1921	CGGAGTCTGTCCTGTACCAAGCTCTGAAATAACCTTAAGAAGGCTTGCMAAGCATTCGACCAA	1980
QY	1981	tgaaaaacatcacagaacaatgacagaagcgggagacacatacttaactgtcgtgtgatatgct	2040
DB	1981	TGAAAAACATCACAGCAATGTACGAGACGGGACACATATCTACATGTGTGTGATTTGCT	2040
QY	2041	atccgctttactcatataaattgtataaagaactctaactgttctccttcaacaacggc	2100
DB	2041	ATCCGCTTACTCATTAATAATGTGTACAAACACTTAACCTGCTTCTTTCAACCAACGCC	2100
QY	2101	aagtcgactttgagaacaactggaagactctaacttcaagatgtgcanaaggaagtgcagc	2160



Oy	363	ggtgcggaagcgagcgagagccgggcccggcgagcgcgagggcgagggcgcgctc	422
Db	9	ggcggtgtgggaagccgcttaccgcggggaagggcctcccgccacggccgacagagcca	68
Oy	423	gtccgcgggaagcgcgaggagcgagcgagcaltgtagatctgctctccgtttccag	482
Db	69	ctctcccccagagaaatagggccctctttatggcatatgctggctgatactttcctctcctg	128
Oy	483	agc---ttcggaatgttttcagtgaaagccaaactccaaactccacttacttgcatagca	539
Db	129	gactctttacaaagaccgccccttaaatgttggcaccacacctcttactttgtaaatgac	188
Oy	540	tctctycaagaagtagtgtgttgcagcaacacggygaactctgtgtccctgcgccgagca	599
Db	189	tctcttgcacacagatgacctttttccagcttccgtggggggtggtggccctgcggcgccg	248
Oy	600	ggatcccccctctgtactctcaatagtagtaccatagcagcgagggagatctacatgtc	659
Db	249	gggtcccccac---ggggcccttgcagtgatgacttgccccaac---ggagcagaca---ctacgacgtg	308
Oy	660	cccggagatccgcaactccaccccaacgagcaactctccaatlttcccttcctccctca	719
Db	309	ccgacatctccggcacagctccagccacacggagcgctgcagcttaccctttccctctc	368
Oy	720	agcttcagtaactatcatcagataaactatattatgcccagctgaaatcctctaagg	779
Db	369	gccttcattagctttatctccagacacaaatgacttacttctgcacccgggaacagctgcggc	428
Oy	780	aaaattagaagtcaggaatgtccacatcaagagctgttttcaagjagccctaatacacgtc	839
Db	429	aagatcccgagcccccacacatccgggtcaaaagcagttttcagggaacccctaacccgtcgg	488
Oy	840	gtggagagacgaagaaacacatgagaggaatgtgtgggtcttcaagtacatccccctc	899
Db	489	gtggagagatcaaaagctacatgcgtggcagacgtggccgtcttcaaggctctcatccccctt	548
Oy	900	lccgttgaagcgatacatcactgtctgcatactggagagaagaacactgtttcactgtctca	959
Db	549	tcagtgacgggaatattgttagcggtgtgttcttggagaaacacacagcttccatcatcca	608
Oy	960	ggaatcaaatcttcacatcacacccacggagggccttgcataataagatgacagaatgaa	1019
Db	609	gaaaacaggtttttttatttaccttaccacgagcgccgtcttacaattcttccacttaccagagag	668
Oy	1020	gaatgattgataaacactacgcctgcatacagcgagcatcagatacccggaagagcgagag	1079
Db	669	gacgcctcttccacacttattccctgcattaccacacacagatgaaagggggagacccggcag	728
Oy	1080	agcaagacggcgcaaaacttttgcataagaccacggaactcgaactcccatcactatgtat	1139
Db	729	agcattatgggacacccctctctgtgtgacagaccctgcgtgaatccatccacacattcttgat	788
Oy	1140	agatttgacatcgcaaaagcagatcgcttggcgcaagcgtgtgtagctgcttgcagaaagcgtc	1199
Db	789	ggctttcacactccacagaaagtgtggcgggcagcaacccttgagactgcttcgacccgctcg	848
Oy	1200	ggagaccctgaagccagattaccgcgtgtcgtgaagagacaatgcccctcgtgaatttcaag	1259
Db	849	ggcttacccttacccccgcacatccctgtgcttgcacagaccctgcgtgaatccacacattcttgat	908
Oy	1260	agatttcaagaagacgcttgagcgagcgctgcattatgaaacatttgcggcttgagtaagc	1319
Db	909	cgcgtgacacaaagcagatcacagggcgtgacattatcagagatttccgagaccgagacagcgcg	968
Oy	1320	agctatgtttgttaaatgtgccaaacagatacgaagactgtcctaagtgatagccgcgtgtac	1379
Db	969	accttacttttgttagtgcacaaacaccttccgttgcggagagacacacagccatccctatg	1028

RESULT 2  
 US-08-348-006B-6  
 Sequence 6, Application US/08348006B  
 Patent No. 5658756  
 GENERAL INFORMATION:  
 APPLICANT: RODAN, GIDEON A.  
 APPLICANT: SCHMIDT, AZREL  
 APPLICANT: RUTLEDGE, SU JANE  
 TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
 TITLE OF INVENTION: TYROSINE PHOSPHATASE  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: J. MARK HAND  
 STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
 CITY: RAMWAY  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07065-0900  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/348,006B  
 FILING DATE:  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/122,032  
 FILING DATE: 14-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HAND, J., MARK  
 REGISTRATION NUMBER: 36,545  
 REFERENCE/DOCKET NUMBER: 189921A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 908-594-3905  
 TELEFAX: 908-594-4720  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6000 base pairs

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 14:46:04 ; Search time 4987.54 Seconds  
(without alignments)  
4999.355 Million cell updates/sec

Title: US-08-956-991-1  
Perfect score: 6604  
Sequence: 1 tgacgagcgccgagcaccg.....gaaatgcccaatatatt 6604

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: gb\_est33: \*  
53: gb\_est34: \*  
54: gb\_est35: \*  
55: gb\_est36: \*  
56: gb\_est37: \*  
57: gb\_est38: \*  
58: gb\_est39: \*  
59: gb\_est40: \*  
60: gb\_est41: \*  
61: gb\_est42: \*  
62: gb\_est43: \*  
63: gb\_est44: \*  
64: gb\_est45: \*  
65: gb\_est46: \*  
66: gb\_est47: \*  
67: gb\_est48: \*  
68: gb\_est49: \*  
69: gb\_est50: \*  
70: gb\_est51: \*  
71: gb\_est52: \*  
72: gb\_est53: \*  
73: gb\_est54: \*  
74: gb\_est55: \*  
75: gb\_est56: \*  
76: gb\_est57: \*  
77: gb\_est58: \*  
78: gb\_est59: \*  
79: gb\_est60: \*  
80: gb\_est61: \*  
81: gb\_est62: \*  
82: gb\_est63: \*  
83: gb\_est64: \*  
84: gb\_est65: \*  
85: gb\_est66: \*  
86: gb\_est67: \*  
87: gb\_est68: \*  
88: gb\_est69: \*  
89: gb\_est70: \*  
90: gb\_est71: \*  
91: gb\_est72: \*  
92: gb\_est73: \*  
93: gb\_est74: \*  
94: gb\_est75: \*  
95: gb\_est76: \*  
96: gb\_est77: \*  
97: gb\_est78: \*  
98: gb\_est79: \*  
99: gb\_est80: \*  
100: gb\_est81: \*  
101: gb\_est82: \*  
102: gb\_est83: \*  
103: gb\_est84: \*  
104: gb\_est85: \*  
105: gb\_est86: \*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	415.8	6.3	439	43	AI243628	AI243628 qb88a03.x
C 2	399.2	6.0	541	64	AW046213	AW046213 UT-M-BH1





\*\*\*\*\*  
MISSED SEQUENCE (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mprrch\_lp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Apr 25 12:30:19 2000; MasPar time 122.36 Seconds  
736.349 Million cell updates/sec  
Tabular output not generated.

Title: >J5-08-956-991-2  
Description: (1-1910) from US08956991A.pep  
Perfect Score: 13516  
Sequence: 1 MWILSLSFQSFANVFSEDL.....KAIGQVSYICLHLENTFC 1910

Scoring table:  
PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 57.841; Variance 116.313; scale 0.497

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	13406	99.2	1896	2	T08851	Down syndrome cell ad	0.00e+00
2	809	8.0	1028	2	I58164	BIG-1 protein - rat	4.31e-127
3	785	5.8	1028	2	A53449	Plasmacytoma-associa	1.80e-122
4	740	5.5	1447	2	A54100	tumor suppressor prot	2.92e-113
5	744	5.5	7962	2	I38346	elastic filin - human	4.63e-114
6	720	5.3	1040	2	A49356	transient axonal glyco	2.85e-109
7	713	5.3	2222	2	T13924	skin protein - fruit f	7.07e-108
8	703	5.2	1036	2	S23283	axonin 1 precursor -	6.90e-106
9	701	5.2	1040	2	A34695	axonal glycoprotein T	1.72e-105
10	707	5.2	1427	2	I51669	tumor suppressor - Af	1.11e-106
11	680	5.0	1259	2	A43425	Bravo/Nr-CAM cell adh	2.55e-101
12	677	5.0	1268	1	A39640	neural cell adhesion	1.00e-100
13	663	4.9	1010	2	JU0094	F11 protein precursor	5.94e-98
14	663	4.9	1091	2	S01998	contactin precursor -	5.94e-98
15	631	4.7	1018	2	A54744	contactin 1 precursor	1.23e-91
16	629	4.7	1018	2	I37246	contactin - human	3.05e-91
17	631	4.7	1020	2	S05944	neural cell surface	1.23e-91
18	631	4.7	1021	2	A57112	contactin precursor -	1.23e-91
19	637	4.7	1259	2	S36126	neural cell adhesion	8.10e-93
20	637	4.7	1260	1	S05479	neural cell adhesion	8.10e-93
21	639	4.7	1651	2	T14160	transmembrane recepto	3.05e-91
22	640	4.7	1698	2	S46216	leukocyte antigen-rel	2.07e-93
23	620	4.6	1257	1	A41060	neural cell adhesion	1.80e-89

24	622	4.6	1894	2	C54689	protein-tyrosine-phos	7.27e-90
25	613	4.5	1018	2	JC4211	neural adhesion prote	4.27e-88
26	609	4.5	2029	1	TDFELK	protein-tyrosine-phos	2.60e-87
27	598	4.4	1897	1	TDRHUK	leukocyte antigen-rel	3.73e-85
28	587	4.3	1323	2	PN0568	connectin 38 - chick	5.31e-83
29	570	4.2	1907	2	S50893	protein-tyrosine-phos	1.10e-79
30	574	4.2	1912	2	A56178	protein-tyrosine-phos	1.83e-80
31	556	4.1	1091	1	IJCJNL	neural cell adhesion	5.86e-77
32	555	4.1	1239	1	A32579	neuroglin - fruit fl	9.17e-77
33	522	3.9	1239	2	JE0099	neural cell adhesion	2.26e-70
34	527	3.9	1070	2	JC4593	protein-tyrosine kin	2.44e-71
35	524	3.9	1088	1	IJXLNL	neural cell adhesion	9.28e-71
36	523	3.9	1092	1	JN0635	neural cell adhesion	1.45e-70
37	521	3.9	1691	1	DS4689	neural cell adhesion	3.52e-70
38	518	3.8	725	2	JE0100	neural cell adhesion	1.33e-69
39	509	3.8	761	1	IJHUNG	neural cell adhesion	7.22e-68
40	498	3.7	853	1	IJBONC	neural cell adhesion	9.37e-66
41	496	3.7	1115	1	IJMSNL	neural cell adhesion	2.27e-65
42	501	3.7	1265	1	A37967	neural cell adhesion	2.49e-66
43	505	3.7	1443	2	I50600	neogenin - chicken (f	4.24e-67
44	504	3.7	1663	2	S46217	protein-tyrosine-phos	6.60e-67
45	490	3.6	1344	2	T14316	rig-1 protein - mouse	3.20e-64

## ALIGNMENTS

RESULT 1  
ENTRY T08851 #type fragment  
TITLE Down syndrome cell adhesion protein 1 - human (fragment)  
ALTERNATE\_NAMES  
ORGANISM  
#formal\_name Homo sapiens #common\_name man  
DATE 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

ACCESSIONS  
REFERENCE T08851  
#authors Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenberg, J.R.  
#submission submitted to the EMBL Data Library, September 1997  
#description DSCAM: A novel member of the immunoglobulin superfamily maps in a down syndrome region and is involved in the development of the nervous system.

#accession T08851  
#status preliminary: translated from GB/EMBL/DBJ  
#molecule\_type mRNA  
#residues 1-1896 #label YAM  
##cross-references EMBL:AF023449; NID:q3169765; PID:q3169766  
##experimental\_source brain; developmental stage: 14 weeks; fetal

## GENETICS

#gene DSCAM  
#map\_position 21q22  
#note derived from alternatively-spliced mRNA

FUNCTION  
#description involved in nervous system development  
KEYWORDS  
#length 1896 #checksum 8680  
SUMMARY

Query Match 99.2%; Score 13406; DB 2; Length 1896;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 1896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


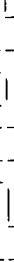



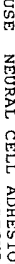
Db	1	VFSEDLHSLYFNASLOEVVFSTGTGLVPCPAAGIPYTLRWYLAATGEELYDVGVGRH	60
Qy	15	VFSFDLHSLYFNASLOEVVFSTGTGLVPCPAAGIPYTLRWYLAATGEELYDVGVGRH	74
Db	61	VHPGTLQIFPPPPSSTLIHNTYCTAENPSGKIRSDQVHKAVLEPTVRYEDQK	120
Qy	75	VHPGTLQIFPPPPSSTLIHNTYCTAENPSGKIRSDQVHKAVLEPTVRYEDQK	134
Db	121	TMKGNVAFKCIIPSSVEAYITVSWEKDVLVSSGRELITSTGALYIKDVONEDGLYN	180
Qy	135	TMKGNVAFKCIIPSSVEAYITVSWEKDVLVSSGRELITSTGALYIKDVONEDGLYN	194
Db	181	YRCITRIRYGTGRQSSAFLFVSDPANSPSLDDGFDRHKMAAGRVLEPCAKJGHPFP	240

```
195 YRCLTIRHRYTGETRQSSARLFVSDPANSAFSLIDGDFHKKAAAGORVELPCALHHPER 254
241 DYRLKDNMBLELSGRFOKTVTGTLLENIRPSSGSGSYCEVSNRYGTAKVIGRLYKQPL 300
255 DYRLKDNMBLELSGRFOKTVTGTLLENIRPSSGSGSYCEVSNRYGTAKVIGRLYKQPL 314
301 KATISPRKVVSSVSGSOVSLSCVTGTEDOLSWYRNGELINPKGNRTIGINENLIMH 360
315 KATISPRKVVSSVSGSOVSLSCVTGTEDOLSWYRNGELINPKGNRTIGINENLIMH 374
361 MWSSDGAYOCFARKDKLSADODYVOVULEDGTPEIISAFSEKVVSPAEPYSLMCNKGTP 420
375 MWSSDGAYOCFARKDKLSADODYVOVULEDGTPEIISAFSEKVVSPAEPYSLMCNKGTP 434
421 LPTITWTLDDDPILKGGSHRISOMITSEGNVSYLNISSQVBDGCVYRCTANNSAGVVL 480
435 LPTITWTLDDDPILKGGSHRISOMITSEGNVSYLNISSQVBDGCVYRCTANNSAGVVL 494
481 YQARIINVRGPASIRPMKNITAIAGROTYIHCRTVIGPYYSIKMKNSNLLPEHNRQVAF 540
495 YQARIINVRGPASIRPMKNITAIAGROTYIHCRTVIGPYYSIKMKNSNLLPEHNRQVAF 554
541 NNGTLKLSDVQKEVEDEGEYCNVLPQPLSTSSVHTVVPFIOFEPFERSIGORVF 600
555 NNGTLKLSDVQKEVEDEGEYCNVLPQPLSTSSVHTVVPFIOFEPFERSIGORVF 614
601 IPCVYVSGDLPIITTTOKODRPIPGSLGVTIIDNTFTSLRISNLSLMHNGNTTCIARNE 660
615 IPCVYVSGDLPIITTTOKODRPIPGSLGVTIIDNTFTSLRISNLSLMHNGNTTCIARNE 674
661 AAAREHOSOLIVRPPKPVVOPRODGIYKAVILNCSAGYEPPTIVMFKSGAGVPOF 720
675 AAAREHOSOLIVRPPKPVVOPRODGIYKAVILNCSAGYEPPTIVMFKSGAGVPOF 734
721 QPILANGRIQVLSNGSLLIKHYVEEDSGYLLCKVSNVDGADVSKMYLTVKIPAMITSTP 780
735 QPILANGRIQVLSNGSLLIKHYVEEDSGYLLCKVSNVDGADVSKMYLTVKIPAMITSTP 794
781 NTTLATOGQKKEKEMCTAHGEKPIITVRWEKEDRIINPEMARIVLSTKVEGEVISTLOILP 840
795 NTTLATOGQKKEKEMCTAHGEKPIITVRWEKEDRIINPEMARIVLSTKVEGEVISTLOILP 854
841 TVRDSGFFSCHAINSNGEDRGIIQLTVQEPDPPEIEIKDVAKRTITLWMTMGFDGNSP 900
855 TVRDSGFFSCHAINSNGEDRGIIQLTVQEPDPPEIEIKDVAKRTITLWMTMGFDGNSP 914
901 ITGYDIECKNKSQSDMSAQRTKDVSPOLNSATIIDIHPSSTYSIRMYAKNRIGSEPSNE 960
915 ITGYDIECKNKSQSDMSAQRTKDVSPOLNSATIIDIHPSSTYSIRMYAKNRIGSEPSNE 974
961 LTIADAEAPRGPPOEYHLEPISSOSIRVTMKAPKXHLONGIINGYOIGRXYSTGNGNFQ 1020
975 LTIADAEAPRGPPOEYHLEPISSOSIRVTMKAPKXHLONGIINGYOIGRXYSTGNGNFQ 1034
1021 FNISVDTSGSEYVYTDNLNKKFTQYGLVQACNRAGTGGSSOEIITTTLEDVPSYPEN 1080
1035 FNISVDTSGSEYVYTDNLNKKFTQYGLVQACNRAGTGGSSOEIITTTLEDVPSYPEN 1094
1081 VOAIATSPESISTSMSTLSKALNGILOGFRVIYMANLMDGELCEIKNITTTQPSLELDG 1140
1095 VOAIATSPESISTSMSTLSKALNGILOGFRVIYMANLMDGELCEIKNITTTQPSLELDG 1154
1141 LEKTYNSIOVLATFRAGDDGYRSEOFIIRTKEDEVPGPAGVKAASASAMVFWMLPPLK 1200
1155 LEKTYNSIOVLATFRAGDDGYRSEOFIIRTKEDEVPGPAGVKAASASAMVFWMLPPLK 1214
1201 LNGIIRKTYVFCSHPYPTVISEFASPDPSFYRIPLNLSRNQOYSVWVAVYTSAGRGSSE 1260
1215 LNGIIRKTYVFCSHPYPTVISEFASPDPSFYRIPLNLSRNQOYSVWVAVYTSAGRGSSE 1274
1261 IITVEPLAKAARILITTSYGTITPMKMDIYLPCKAVGDPSP7 VMMMDNSNGTSPSLVTIDG 1320
```

```
QY 1275 IITVEPLAKAARILITTSYGTITPMKMDIYLPCKAVGDPSP7VMMMDNSNGTSPSLVTIDG 1334
Db 1321 RRSITFNGSFIITRTVKAEDSGYISCIANNMGSDEITLNLQVOYPPPOPLJYSKTITSS 1380
QY 1335 RRSITFNGSFIITRTVKAEDSGYISCIANNMGSDEITLNLQVOYPPPOPLJYSKTITSS 1394
Db 1381 IITLWLPDNGSGSSIRGIILOYSEDNSEOWGSPISPSERSYRLNENKCGTWKFKTLTAQ 1440
QY 1395 IITLWLPDNGSGSSIRGIILOYSEDNSEOWGSPISPSERSYRLNENKCGTWKFKTLTAQ 1454
Db 1441 NGVGRIRSEIIEAKTLCKEPOFSKEOLFASINTTRVRLNLIGMNDGGCPITSFTLEYR 1500
QY 1455 NGVGRIRSEIIEAKTLCKEPOFSKEOLFASINTTRVRLNLIGMNDGGCPITSFTLEYR 1514
Db 1501 PEGTIVWTTAORTSLSKSYILYDLOEATWYELQWRVNCNSAGACAKQANFATLNDGSTIP 1560
QY 1515 PEGTIVWTTAORTSLSKSYILYDLOEATWYELQWRVNCNSAGACAKQANFATLNDGSTIP 1574
Db 1561 PLIKSVQNEEGLTNEGKMLVTISCIIVGVLLFVLLVYRRRREORLRLRDAKSL 1620
QY 1575 PLIKSVQNEEGLTNEGKMLVTISCIIVGVLLFVLLVYRRRREORLRLRDAKSL 1634
Db 1621 AEMLSKNTRTSDTLSKOQOTLRMHIDIPRAQLLIEERDWTETIDDSYSTVLLDADGFA 1680
QY 1635 AEMLSKNTRTSDTLSKOQOTLRMHIDIPRAQLLIEERDWTETIDDSYSTVLLDADGFA 1694
Db 1681 AKOKSLTHTVHVSQOATGPLVDSDARPNTTRAKAGPAPARNRYASQWTLNR 1740
QY 1695 AKOKSLTHTVHVSQOATGPLVDSDARPNTTRAKAGPAPARNRYASQWTLNR 1754
Db 1741 PAPTISATHTLTDMRLPFPRAAGSYDKESDSYSVSPQDTRARRSNVSTESASTYEEL 1800
QY 1755 PAPTISATHTLTDMRLPFPRAAGSYDKESDSYSVSPQDTRARRSNVSTESASTYEEL 1814
Db 1801 ARAYEHA WEDOLRAKATTIECFISDTSSQULAGTNEYDLSITSPSSGICRTAS 1860
QY 1815 ARAYEHA WEDOLRAKATTIECFISDTSSQULAGTNEYDLSITSPSSGICRTAS 1874
Db 1861 PPKPODGGRYMMAVPAKAIQOVTSYICLHTEMTFC 1896
QY 1875 PPKPODGGRYMMAVPAKAIQOVTSYICLHTEMTFC 1910
```

2

RESULT	ENTRY	#	type	complete
ENTRY	158164	1	BIG-1 protein	- rat
TITLE			formal_name	Kallus norvegicus
ORGANISM			common_name	Norway rat
DATE			sequence_revision	26-Jul-1996
			text_change	24-Sep-1996
ACCESSIONS	158164			
REFERENCE	158164			
#authors				Yoshihara, Y.; Kawasaki, M.; Tani, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K.
#journal				Neuron (1994) 13:415-426
#title				BIG-1: a new TAG-1/F3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity.
#cross-references	MUID:94338697			
#accession	158164			
#status				preliminary; translated from GB/EMBL/DBJ
GENETICS				
#molecule_type	mRNA			
#residues	1-1028			#label RES
#cross-references	EMBL:U11031; NID:g563132; PIDN:AAA65607.1; PID:g563133			
CLASSIFICATION				
#gene	BIG-1			
#superfamily	contactin; fibronectin type III repeat homology;			
length	1028			#molecular_weight 112788
checksum	5866			#checksum 5866
Query Match				
Best local similarity	24.7%			Pred. No. 4,31e-127;
Matches	250;			Conservative 249; Mismatches 447; Indels 66; Gaps 54;





```

|||||
OY 255 DYRWLMDNMPLELSGRFQKTVTGLLENIIRPSDSCSYCEVSNRGTAKVIGRLTVKQPL 314
Db 301 KATISPRKXSSVGSQVSLSCSVTGTDEDELISWYRNGELNPGKNVTRITGTHHEMLMDH 360
OY 315 KATISPRKXSSVGSQVSLSCSVTGTDEDELISWYRNGELNPGKNVTRITGTHHEMLMDH 374
Db 361 MKSDGAYQCEVFRKDKLSAODYVQVVLDEDTPKIISAFSEKVVSPAPVSLMCVNGKTP 420
OY 375 MKSDGAYQCEVFRKDKLSAODYVQVVLDEDTPKIISAFSEKVVSPAPVSLMCVNGKTP 434
Db 421 LPTITWTDLDDPILKGGSHRISQMITSEGNVSYLNISSSQV?CGVYRGTANNAGVYL 480
OY 435 LPTITWTDLDDPILKGGSHRISQMITSEGNVSYLNISSSQV?CGVYRGTANNAGVYL 494
Db 481 YQARINVRGPASTIRPKNTITAGRDYTHCRVIGYPRYSIKWYNSNLPRNHQVAF 540
OY 495 YQARINVRGPASTIRPKNTITAGRDYTHCRVIGYPRYSIKWYNSNLPRNHQVAF 554
Db 541 NNGTLKLSQKVEDEGEYTCNVLPOLSTSQSVHVTYKVPFIQPEEPFISIGRVF 600
OY 555 NNGTLKLSQKVEDEGEYTCNVLPOLSTSQSVHVTYKVPFIQPEEPFISIGRVF 614
Db 601 PCVWVSGDLPTITWQKGRPIPGSLGYTIDNIDFTSLRLSNLSLHNGNYTICIANE 660
OY 615 PCVWVSGDLPTITWQKGRPIPGSLGYTIDNIDFTSLRLSNLSLHNGNYTICIANE 674
Db 661 AAAVEHOSOLIVRPKPEVQPRDDGIGKAVILNCSEGEYPTIYWKFSKAGVQF 720
OY 675 AAAVEHOSOLIVRPKPEVQPRDDGIGKAVILNCSEGEYPTIYWKFSKAGVQF 734
Db 721 QPIALNGRIQVLSNGSLIKHVEEDSGYLLCKVSNVGVADVSKSMYLTVKIPAMITSY 780
OY 735 QPIALNGRIQVLSNGSLIKHVEEDSGYLLCKVSNVGVADVSKSMYLTVKIPAMITSY 794
Db 781 NNTLATQGGKMKSTABEKEPIYVRMEKEDIINPEAK:IVSTKEVEEYISTLOLTP 840
OY 795 NNTLATQGGKMKSTABEKEPIYVRMEKEDIINPEAK:IVSTKEVEEYISTLOLTP 854
Db 841 TVREDSGFSCSHAINSYGEDRGIIQLTVOEPPDEIEIKDKARTIILRMTMGFDGNSP 900
OY 855 TVREDSGFSCSHAINSYGEDRGIIQLTVOEPPDEIEIKDKARTIILRMTMGFDGNSP 914
Db 901 ITGYDIECKNKSQSDSAQRTKDVSPOLNSATIIDIHPSSTYSIMVAKNRIGKSEPSNE 960
OY 915 ITGYDIECKNKSQSDSAQRTKDVSPOLNSATIIDIHPSSTYSIMVAKNRIGKSEPSNE 974
Db 961 LTTTADDEAPDPDPPOEVHLEPISSOSIRVTWKAPKKHLQNGIIRGYOIGREYSTGGMFQ 1020
OY 975 LTTTADDEAPDPDPPOEVHLEPISSOSIRVTWKAPKKHLQNGIIRGYOIGREYSTGGMFQ 1034
Db 1021 FNIIISVDTSGDSEVYTLDLNKFQYGLVQACNAGTGPSSQELITTTLEDVPSYPEN 1080
OY 1035 FNIIISVDTSGDSEVYTLDLNKFQYGLVQACNAGTGPSSQELITTTLEDVPSYPEN 1094
Db 1081 VOALATSPESISISWSTLSKEALNGILOGFRVIYVWNLMDGELGKINNTTQPSLELDG 1140
OY 1095 VOALATSPESISISWSTLSKEALNGILOGFRVIYVWNLMDGELGKINNTTQPSLELDG 1154
Db 1141 LEKYNYSIOVLAFTRAGGVSEIOIFTRTKEDVGPAGVAKAAASASMFVSLPLPK 1200
OY 1155 LEKYNYSIOVLAFTRAGGVSEIOIFTRTKEDVGPAGVAKAAASASMFVSLPLPK 1214
Db 1201 LKNGIRKTYVFCSHRYPTVISEFEASPDSEYRIPNLSRN.QYSWVAVVTSAGKNSSE 1260
OY 1215 LKNGIRKTYVFCSHRYPTVISEFEASPDSEYRIPNLSRNQYSWVAVVTSAGKNSSE 1274
Db 1261 ITTVEPLAARPARILTFSGTITVTPMKKDIIVLPCKAVGDPSPAVKMKKSNCTPSLVTIDG 1320
OY 1275 ITTVEPLAARPARILTFSGTITVTPMKKDIIVLPCKAVGDPSPAVKMKKSNCTPSLVTIDG 1334
Db 1321 RRSIFSNGSEFIIRTVKAEDSGYSCIANNNMGSDIILNLQOVPPDPRLTVSKTSSS 1380
|||||

```

```

OY 1335 RRSIFSNGSEFIIRTVKAEDSGYSCIANNNMGSDIILNLQOVPPDPRLTVSKTSSS 1394
Db 1381 ITLSMLPGDNGSSIRGCIYLIQYSEDNSEDQMGSPFIPSPERSRLEMLKCGTYWKTFLTAQ 1440
OY 1395 ITLSMLPGDNGSSIRGCIYLIQYSEDNSEDQMGSPFIPSPERSRLEMLKCGTYWKTFLTAQ 1454
Db 1441 NGVOPGRISFIIIAKTLGKEPOFSKEQELFASINTFRVRLNLIGMWDGCPITSFLEVR 1500
OY 1455 NGVOPGRISFIIIAKTLGKEPOFSKEQELFASINTFRVRLNLIGMWDGCPITSFLEVR 1514
Db 1501 PFGTTWTTAORTSLSKSYLLYDLEATWYELQMRVCSAGCAEQANFATLNDGSTIP 1560
OY 1515 PFGTTWTTAORTSLSKSYLLYDLEATWYELQMRVCSAGCAEQANFATLNDGSTIP 1574
Db 1561 PLTKSVQNEEGTLTNEGKMLNVTISCLLVGVLFLVLLVRRRRRERLRLDASL 1620
OY 1575 PLTKSVQNEEGTLTNEGKMLNVTISCLLVGVLFLVLLVRRRRRERLRLDASL 1634
Db 1621 AEMLSKNTRTSDTLKQOQTLRMHIDIPRAQLIEERDTMETIDRSTVLLTDADFGFA 1680
OY 1635 AEMLSKNTRTSDTLKQOQTLRMHIDIPRAQLIEERDTMETIDRSTVLLTDADFGFA 1694
Db 1681 AKOKSLTVHTVHYQSVQATGCLVDVSDARPGTNETTRNAKAGPTANRYASQWTLNR 1740
OY 1695 AKOKSLTVHTVHYQSVQATGCLVDVSDARPGTNETTRNAKAGPTANRYASQWTLNR 1754
Db 1741 PHPTISAHITLTDWRLPTPRAAGSVDESDSYSVSPQDIDARSSMTSESASTYEEL 1800
OY 1755 PHPTISAHITLTDWRLPTPRAAGSVDESDSYSVSPQDIDARSSMTSESASTYEEL 1814
Db 1801 ARAYEHAKMEQLRHAKFTITTECFISDTSSEDLTAGTNETYSLTSTI SESEIGRFTAS 1860
OY 1815 ARAYEHAKMEQLRHAKFTITTECFISDTSSEDLTAGTNETYSLTSTI SESEIGRFTAS 1874
Db 1861 PPKPODGGRYMNAVPAKATGQVTSYICLTLTLEWTCF 1896
OY 1875 PPKPODGGRYMNAVPAKATGQVTSYICLTLTLEWTCF 1910

RESULT 2
ID 060469 PRELIMINARY: PRT: 1571 AA.
AC 060469:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE.
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA YAMAKANA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RA Submitted (sep-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF023450; AAC17967.1;
DR PRAW: PF00041; fn3: 6.
DR PRAW: PF00047; ig: 8.
DR PRINTS: PF00014; FNTYPEP11.
SQ SEQUENCE 1571 AA: 5F8C77D1 CRC32:

Query Match 82.3%; Score 11119; DA 4; Length 1571;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1561; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:09 ; Search time 383.18 Seconds  
(without alignments)  
13.059 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20

Sequence: 1 ccaggtctcaagaagcagcagc 20

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 23170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance -v have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	V31983	Human Down syndrom
2	20	100.0	6413	V31988	Human Down syndrom
3	20	100.0	6604	V31981	Human Down syndrom
4	18.4	92.0	2173	V31987	Mouse Down syndrom
5	16.8	84.0	187	T22739	Human gene signatu
6	16.4	82.0	12923	N90338	Sequence of human
7	15.8	79.0	5406	V52299	Streptococcus pneu
8	15.4	77.0	4646	X20577	Poly nucleotide seq
9	15.2	76.0	3417	T39622	Human DNA ligase I
10	15.2	76.0	3417	V82491	Human DNA ligase I
11	15.2	76.0	8298	1	PI90 protein gene.
12	15.2	76.0	12752	1	Sequence which cor
13	15.2	76.0	110000	T58840_1	Continuation (2 of
14	14.8	74.0	2071	V46494	Interleukin-1 beta
15	14.8	74.0	2975	X03792	Human semaphorin E
16	14.8	74.0	8450	T74994	Plasmid pFU1001 co
17	14.8	74.0	8535	1	Galv SEB10 genome.
18	14.8	74.0	9326	1	DNA complementary
19	14.4	72.0	709	X07563	Homo sapiens eosin
20	14.4	72.0	1185	V43608	Human secreted pro
21	14.4	72.0	1537	1	Human secreted pro
22	14.4	72.0	1605	V59687	Human secreted pro
23	14.4	72.0	1871	1	Guinea pig PAR rec
24	14.4	72.0	2132	1	Interferon-pseudo
25	14.4	72.0	2132	1	Arabidopsis CER1 g
26	14.4	72.0	2530	1	Mouse beta-3 integ
27	14.4	72.0	3517	1	Mouse truncated be
28	14.4	72.0	3534	1	H. pylori cytoplas
29	14.4	72.0	3549	1	H. pylori cytoplas
30	14.4	72.0	3648	1	120-128 kilodalton
31	14.4	72.0	3648	1	Helicobacter pylori
32	14.4	72.0	4149	1	Helicobacter pylori
33	14.4	72.0	4821	1	Tag A antigen of H
34	14.4	72.0	4821	1	H. pylori tagA ant

35	14.4	72.0	4821	1	T14051	Helicobacter pylori
36	14.4	72.0	5925	1	Q48733	CAI gene. Recombin
37	14.4	72.0	19932	1	T46159	CagI locus. Helico
38	14.2	71.0	288	1	T60756	HIV-1 strain YBF30
39	14.2	71.0	560	1	T59027	PDD16b left juncti
40	14.2	71.0	580	1	T16308	Pea H74 CDNA parti
41	14.2	71.0	600	1	V21212	Human type II GNRH
42	14.2	71.0	600	1	V26007	Bacteriophage DNA
43	14.2	71.0	874	1	V59097	Clathrin-associat
44	14.2	71.0	1101	1	V59654	Human secreted pro
45	14.2	71.0	1422	1	O63870	Feline T cell prot

## ALIGNMENTS

RESULT 1	
ID V31983	V31983 standard; cDNA; 20 BP.
AC V31983	28-SEP-1998 (first entry)
DE Human Down syndrome-cell adhesion molecule DS-CAM primer B9-131F.	
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; PCR; primer: ss.	
OS Synthetic.	
OS Homo sapiens.	
PN W09817795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; U19547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI Korenberg JR;	
DR MPI: 98-271791/24.	
PT New isolated Down's Syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Example 5; Page 79; 109pp; English.	
CC Forward primer B9-131F and reverse primer B9-131R (see V31984)	
CC were used in RT-PCR assays of cDNA libraries of various human	
CC tissues to determine human Down syndrome-cell adhesion molecule	
CC DS-CAM (see V31981 and V31988) expression. The results	
CC demonstrated expression of human DS-CAM mRNA in foetal and adult	
CC brain, and foetal kidney. In addition, a breast carcinoma cell	
CC line showed expression of human DS-CAM mRNA. DS-CAM polypeptides	
CC (see W42086-87) are associated with developmental and neurological	
CC processes. The polypeptides and nucleic acids are used to develop	
CC and neurological abnormalities.	
SO Sequence 20 BP: 5 A; 5 C; 6 G; 3 T;	
Query Match	100.0%; Score 20; DB 1; Length 20;
Best Local Similarity	100.0%; Pred. No. 0.41; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0;	
QY 1 ccaggtctcaagaagcagc 20	
DB 1 CCAGTCTCAAGAGCAGCAG 20	
RESULT 2	
ID V31988	V31988 standard; cDNA; 6413 BP.
AC V31988	28-SEP-1998 (first entry)
DE Human Down syndrome-cell adhesion molecule DS-CAM2 cDNA.	
DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; ds; ss.	
OS Homo sapiens.	

FH Key Location/Qualifiers  
 FT CDS 453..5168  
 FT /\*tag= a  
 PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.  
 PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR P-PSDB: W42087.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to  
 develop products for detection, diagnosis and therapy of  
 developmental and neurological abnormalities  
 PS Claim 3: Page 83-90: 109pp: English.  
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule  
 DS-CAM2 (see W42087), an extracellular, soluble protein belonging  
 to a novel subclass of the Ig superfamily with highest homology to  
 neural cell adhesion molecules. A modified direct cDNA selection  
 technique was applied to bacterial and pl artificial chromosomes  
 between ETS2 and MX1 by using cDNA from trisomy 21 human foetal  
 brain. A unique cDNA fragment, designated E51 (see V31982), was  
 obtained and used to screen a trisomy 21 human foetal brain (14-wk  
 gestation) cDNA library. Further clones were obtained by exon  
 trapping, and the complete DS-CAM2 cDNA sequence was identified.  
 CC DS-CAM2 is a splice variant of membrane-bound DS-CAM1 (see W42086)  
 and is deleted for the entire transmembrane domain. The DS-CAM  
 gene spans 900-1200 kb of genomic DNA and maps at chromosome  
 21q22.2-22.3. The invention also provides murine DS-CAM partial  
 sequences (see V31985-87), expression vectors and host cells,  
 transgenic animals, antisense oligonucleotides, and primers useful  
 for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are  
 associated with developmental and neurological processes. They can  
 be used in e.g. neural prosthetic devices used in entubulation  
 CC methods of repairing (regenerating) damaged or severed peripheral  
 nerves. The products can also be used in detection, diagnosis and  
 therapy of developmental and neurological abnormalities such as  
 Down syndrome, mental retardation, holoprosencephaly, agenesis of  
 the corpus callosum, or schizencephaly. Antisense oligonucleotides  
 are used for inhibiting translation of mRNA. 1707 G; 1292 T;  
 SQ Sequence 6413 BP; 1633 A; 1781 C;

Query Match 100.0%; Score 20; DB 1; Length 6413;  
 Best Local Similarity 100.0%; Pred. No. 0.62;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagttctcaagagcagc 20  
 DB 4877 CCAGTTCTCAAGAGCAGCAGG 4896  
 |||||||||||||||||||

RESULT 3  
 ID V31981 standard; cDNA: 6604 BP.  
 AC V31981:  
 DT 28-SEP-1998 (first entry)  
 DE Human Down syndrome-cell adhesion molecule; neural cell;  
 KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;  
 KW signal transduction; trisomy 21; mental retardation;  
 KW holoprosencephaly; corpus callosum agenesis;  
 KW schizencephaly; diagnosis; assay; human; ds; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 453..6185  
 FT /\*tag= a  
 FT sig\_peptide 452..521  
 FT /\*tag= b  
 FT mat\_peptide 522..6162  
 FT /\*tag= c  
 PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.

PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 DR P-PSDB: W42086.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to  
 develop products for detection, diagnosis and therapy of  
 developmental and neurological abnormalities  
 PS Claim 3: Page 65-72: 109pp: English.  
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule  
 DS-CAM1 (see W42086), a cell surface glycoprotein belonging to a  
 novel subclass of the Ig superfamily with highest homology to  
 neural cell adhesion molecules. A modified direct cDNA selection  
 technique was applied to bacterial and pl artificial chromosomes  
 between ETS2 and MX1 by using cDNA from trisomy 21 human foetal  
 brain. A unique cDNA fragment, designated E51 (see V31982), was  
 obtained and used to screen a trisomy 21 human foetal brain (14-wk  
 gestation) cDNA library. Further clones were obtained by exon  
 trapping. The complete DS-CAM1 cDNA sequence was deduced from  
 overlapping clones. A splice variant cDNA (see V31988), encoding  
 non-membrane bound DS-CAM2 (see W42087), was also identified. The  
 DS-CAM gene spans 900-1200 kb of genomic DNA and maps at chromosome  
 21q22.2-22.3. The invention also provides murine DS-CAM partial  
 sequences (see V31985-87), expression vectors and host cells,  
 transgenic animals, antisense oligonucleotides, and primers useful  
 for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are  
 associated with developmental and neurological processes. They can  
 be used in e.g. neural prosthetic devices used in entubulation  
 CC methods of repairing (regenerating) damaged or severed peripheral  
 nerves. The products can also be used in detection, diagnosis and  
 therapy of developmental and neurological abnormalities such as  
 Down syndrome, mental retardation, holoprosencephaly, agenesis of  
 the corpus callosum, or schizencephaly. Antisense oligonucleotides  
 are used for inhibiting translation of mRNA. 1827 C; 1768 G; 1336 T;  
 SQ Sequence 6604 BP; 1673 A;

Query Match 100.0%; Score 20; DB 1; Length 6604;  
 Best Local Similarity 100.0%; Pred. No. 0.62;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagttctcaagagcagc 20  
 DB 4877 CCAGTTCTCAAGAGCAGCAGG 4896  
 |||||||||||||||||||

RESULT 4  
 ID V31987 standard; cDNA: 2174 BP.  
 AC V31987:  
 DT 28-SEP-1998 (first entry)  
 DE Mouse Down syndrome-cell adhesion molecule; DS-CAM 3' cDNA.  
 KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;  
 KW signal transduction; trisomy 21; mental retardation;  
 KW holoprosencephaly; corpus callosum agenesis;  
 KW schizencephaly; diagnosis; assay; mouse; ds; ss.  
 OS Mus sp.  
 PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.  
 PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to  
 develop products for detection, diagnosis and therapy of  
 developmental and neurological abnormalities  
 PS Claim 2: Page 81-83: 109pp: English.  
 CC This cDNA sequence comprises the 3' region of a cDNA clone  
 for murine Down syndrome-cell adhesion molecule (DS-CAM), a  
 member of a novel subclass of the Ig superfamily with homology to  
 neural cell adhesion molecules. The middle region (see V31986)  
 and 5' region (see V31985) of the clone are also provided. The



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:19 ; Search time 216.06 Seconds  
(without alignments)  
11.082 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20  
Sequence: 1 ccaggtctcaaggagcagg 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents.NA:\*  
1: /cgcn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgcn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgcn2\_6/ptodata/1/ina/5C.COMB.seq:\*  
4: /cgcn2\_6/ptodata/1/ina/5D.COMB.seq:\*  
5: /cgcn2\_6/ptodata/1/ina/5E.COMB.seq:\*  
6: /cgcn2\_6/ptodata/1/ina/5F.COMB.seq:\*  
7: /cgcn2\_6/ptodata/1/ina/5G.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	15.2	76.0	3417	3	US-08-464-402-1 Sequence 1, Appl 1
2	15.2	76.0	8298	6	PCT-US93-03076-1 Sequence 1, Appl 1
3	15.2	76.0	12752	3	US-08-459-146-1 Sequence 1, Appl 1
4	15.2	76.0	12752	3	US-08-459-065-1 Sequence 1, Appl 1
5	14.4	72.0	1314	1	US-07-662-005A-15 Sequence 15, Appl 1
6	14.4	72.0	3648	2	US-08-053-614-1 Sequence 1, Appl 1
7	14.4	72.0	3648	2	US-08-316-397B-1 Sequence 1, Appl 1
8	14.4	72.0	3648	2	US-09-034-306-1 Sequence 1, Appl 1
9	14.4	72.0	3648	6	PCT-US93-09782-1 Sequence 1, Appl 1
10	14.4	72.0	4821	2	US-08-053-614-3 Sequence 3, Appl 1
11	14.4	72.0	4821	2	US-08-316-397B-3 Sequence 3, Appl 1
12	14.4	72.0	4821	2	US-09-034-306-3 Sequence 3, Appl 1
13	14.4	72.0	4821	6	PCT-US93-09782-3 Sequence 3, Appl 1
14	14.4	72.0	10299	4	US-08-477-451-1 Sequence 1, Appl 1
15	14.4	72.0	10299	4	US-08-477-451-1 Sequence 5, Appl 1
16	14.4	72.0	19932	4	US-08-477-451-25 Sequence 25, Appl 1
17	14.4	72.0	460	4	US-08-487-727A-1 Sequence 1, Appl 1
18	14.2	71.0	580	2	US-08-272-255-19 Sequence 19, Appl 1
19	14.2	71.0	580	6	PCT-US95-08565-19 Sequence 19, Appl 1
20	14.2	71.0	874	3	US-08-850-119-2 Sequence 2, Appl 1
21	14.2	71.0	911	2	US-08-745-603-1 Sequence 1, Appl 1
22	14.2	71.0	1569	4	US-08-923-772-1 Sequence 1, Appl 1
23	14.2	71.0	1617	4	US-08-324-483-1 Sequence 1, Appl 1
24	14.2	71.0	1686	1	US-08-324-483-1 Sequence 1, Appl 1
25	14.2	71.0	1747	3	US-08-522-421-5 Sequence 5, Appl 1
26	14.2	71.0	1812	4	US-08-735-041A-3 Sequence 3, Appl 1
27	14.2	71.0	2103	4	US-08-735-041A-5 Sequence 5, Appl 1

28	14.2	71.0	2103	4	US-08-735-041A-7 Sequence 7, Appl 1
29	14.2	71.0	2559	3	US-08-886-152-4 Sequence 4, Appl 1
30	14.2	71.0	2692	1	US-07-932-454A-2 Sequence 2, Appl 1
31	14.2	71.0	3171	6	PCT-US95-09261-1 Sequence 1, Appl 1
32	14.2	71.0	3215	2	US-08-426-627-1 Sequence 1, Appl 1
33	14.2	71.0	4104	1	US-07-998-003A-94 Sequence 94, Appl 1
34	14.2	71.0	4104	1	US-08-453-274B-94 Sequence 94, Appl 1
35	14.2	71.0	4104	2	US-08-453-695A-94 Sequence 94, Appl 1
36	14.2	71.0	4104	2	US-08-765-161A-94 Sequence 94, Appl 1
37	14.2	71.0	4104	3	US-08-453-702A-94 Sequence 94, Appl 1
38	14.2	71.0	4104	6	PCT-US93-12588-94 Sequence 94, Appl 1
39	14.2	71.0	4104	6	PCT-US95-08071-94 Sequence 94, Appl 1
40	14.2	71.0	4650	1	US-07-998-003A-102 Sequence 102, Appl 1
41	14.2	71.0	4650	1	US-08-453-274B-102 Sequence 102, Appl 1
42	14.2	71.0	4650	2	US-08-453-695A-102 Sequence 102, Appl 1
43	14.2	71.0	4650	2	US-08-268-161A-102 Sequence 102, Appl 1
44	14.2	71.0	4650	3	US-08-453-702A-102 Sequence 102, Appl 1
45	14.2	71.0	4650	6	PCT-US93-12588-102 Sequence 102, Appl 1

## ALIGNMENTS

RESULT 1  
US-08-464-402-1  
Sequence 1, Application US/08464402  
Patent No. 5858705  
GENERAL INFORMATION:  
APPLICANT: WEI, ET AL.  
TITLE OF INVENTION: Human DNA Ligase III  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GIUFFILIAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INC DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,402  
FILING DATE: June 5, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03939  
FILING DATE: 31 MAR 95  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3417 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA  
US-08-464-402-1  
Query Match 76.0%; Score 15.2; DB 3; Length 3417;  
Best Local Similarity 85.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 3; Indels 0;  
Gaps 0;  
OY 1 ccaggtctcaaggagcagg 20

||||| || ||||| |||  
Db 2559 CCACTTCTCAAGCAGCAGC 2578

RESULT 2  
PCT-US93-03076-1  
; Sequence 1, Application PC/TUS9303076  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; TITLE OF INVENTION: GAP-Associated Protein p190 and  
; TITLE OF INVENTION: Transduction  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: 2 Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/03076  
; FILING DATE: 19930331  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WH192-03A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8298 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 731..5272  
; PCT-US93-03076-1

Query Match 76.0%; Score 15.2; DB 6; Length 8298;  
Best Local Similarity 85.0%; Pred. No. 47;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccagttctcaagcagcagc 20  
Db 3763 CCACTTCTCAATGAGCTGG 3782

RESULT 3  
US-08-459-146-1  
; Sequence 1, Application US/08459146  
; Patent No. 5886405  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Gil Ho  
; APPLICANT: Nuss, Donald Lee  
; TITLE OF INVENTION: Genetically Engineered Transmissible  
; TITLE OF INVENTION: Hypovirulence  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; COUNTRY: U.S.A.

ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,146  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/832,117  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roseman, Catherine R  
; REGISTRATION NUMBER: 74,240  
; REFERENCE/DOCKET NUMBER: 8589  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-6208  
; TELEFAX: (201) 235-3500  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12752 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Endothia parasitica (Cryptomycetozoa)  
; ORGANISM: parasitica  
; STRAIN: EP713  
US-08-459-146-1

Query Match 76.0%; Score 15.2; DB 3; Length 12752;  
Best Local Similarity 85.0%; Pred. No. 51;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ccagttctcaagcagcagc 20  
Db 12192 CCAACTGCGCAAGCAGCAGC 12211

RESULT 4  
US-08-459-065-1  
; Sequence 1, Application US/08459065  
; Patent No. 5882642  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Gil Ho  
; APPLICANT: Nuss, Donald Lee  
; TITLE OF INVENTION: Genetically Engineered Transmissible  
; TITLE OF INVENTION: Hypovirulence  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,065  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/832,117

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:36 ; Search time 4987.54 Seconds  
(without alignments)  
15.140 Million cell updates/sec

Title: US-08-956-991-5

Perfect score: 20  
Sequence: 1 ccaagttccaaagagagcagc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database :

## EST:

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est33:\*  
53: gb\_est34:\*  
54: gb\_est35:\*  
55: gb\_est36:\*  
56: gb\_est37:\*  
57: gb\_est38:\*  
58: gb\_est39:\*  
59: gb\_est40:\*  
60: gb\_est41:\*  
61: gb\_est42:\*  
62: gb\_est43:\*  
63: gb\_est44:\*  
64: gb\_est45:\*  
65: gb\_est46:\*  
66: gb\_est47:\*  
67: gb\_est48:\*  
68: gb\_est49:\*  
69: gb\_est50:\*  
70: gb\_est51:\*  
71: gb\_est52:\*  
72: gb\_est53:\*  
73: gb\_est54:\*  
74: gb\_est55:\*  
75: gb\_est56:\*  
76: gb\_est57:\*  
77: gb\_est58:\*  
78: gb\_est59:\*  
79: gb\_est60:\*  
80: gb\_est61:\*  
81: gb\_est62:\*  
82: gb\_est63:\*  
83: gb\_est64:\*  
84: gb\_est65:\*  
85: gb\_est66:\*  
86: gb\_est67:\*  
87: gb\_est68:\*  
88: gb\_est69:\*  
89: gb\_est70:\*  
90: gb\_est71:\*  
91: gb\_est72:\*  
92: gb\_est73:\*  
93: gb\_est74:\*  
94: gb\_est75:\*  
95: gb\_est76:\*  
96: gb\_est77:\*  
97: gb\_est78:\*  
98: gb\_est79:\*  
99: gb\_est80:\*  
100: gb\_est81:\*  
101: gb\_est82:\*  
102: gb\_est83:\*  
103: gb\_est84:\*  
104: gb\_est85:\*  
105: gb\_est86:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.8	84.0	177	74	AM176762
2	16.8	84.0	319	61	AV158019

Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC06c2-RC0-CT0094-200899-002-B12&c3=1999-08-20&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 177.

## FEATURES

Location/Qualifiers

1..177  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0094"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

45 a 38 c 62 g 32 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 74; Length 177;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY

1 ccaggtctcaaggagcagg 20

## DB

37 CCAGGTCTCAAGGAGCAGC 56

## RESULT 2

AV1 8019 319 bp mRNA EST 07-JUL-1999  
AV1 3019 Mus musculus head C57BL/6J 12-day embryo Mus musculus cDNA  
clone 3010012H19, mRNA sequence.

## ACCESSION

AV18019.1 GI:5401654

## KEYWORDS

EST

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 319)

## AUTHORS

Carrincci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

## TITLE

Unpublished (1999)

## JOURNAL

On May 18, 1998

## COMMENT

This sequence version replaced gi:3137650.

## CONTACT

Chie Owa

## GENOME

Science Laboratory

## RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

## TEL

81-298-36-9145

## FAX

81-298-36-9098

## EMAIL

genome-res@cc.riken.go.jp

## THERMOSTABILIZATION

Thermolabile enzymes by

## THERMOSTABILIZATION

and its application for the synthesis of full length cDNA

## TRANSCRIPTIONAL

sequencing: A method for DNA sequencing using RNA

## POLYMERASE (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.riken.go.jp>) for further details.

## LOCATION/QUALIFIERS

1..319

## ORGANISM

"Mus musculus"

## STRAIN

"C57BL/6J"

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
1	AM176762	177 bp mRNA	EST	16-NOV-1999						
2	AM176762	177 bp mRNA	EST	16-NOV-1999						
3	AM176762	177 bp mRNA	EST	16-NOV-1999						
4	AM176762	177 bp mRNA	EST	16-NOV-1999						
5	AM176762	177 bp mRNA	EST	16-NOV-1999						
6	AM176762	177 bp mRNA	EST	16-NOV-1999						
7	AM176762	177 bp mRNA	EST	16-NOV-1999						
8	AM176762	177 bp mRNA	EST	16-NOV-1999						
9	AM176762	177 bp mRNA	EST	16-NOV-1999						
10	AM176762	177 bp mRNA	EST	16-NOV-1999						
11	AM176762	177 bp mRNA	EST	16-NOV-1999						
12	AM176762	177 bp mRNA	EST	16-NOV-1999						
13	AM176762	177 bp mRNA	EST	16-NOV-1999						
14	AM176762	177 bp mRNA	EST	16-NOV-1999						
15	AM176762	177 bp mRNA	EST	16-NOV-1999						
16	AM176762	177 bp mRNA	EST	16-NOV-1999						
17	AM176762	177 bp mRNA	EST	16-NOV-1999						
18	AM176762	177 bp mRNA	EST	16-NOV-1999						
19	AM176762	177 bp mRNA	EST	16-NOV-1999						
20	AM176762	177 bp mRNA	EST	16-NOV-1999						
21	AM176762	177 bp mRNA	EST	16-NOV-1999						
22	AM176762	177 bp mRNA	EST	16-NOV-1999						
23	AM176762	177 bp mRNA	EST	16-NOV-1999						
24	AM176762	177 bp mRNA	EST	16-NOV-1999						
25	AM176762	177 bp mRNA	EST	16-NOV-1999						
26	AM176762	177 bp mRNA	EST	16-NOV-1999						
27	AM176762	177 bp mRNA	EST	16-NOV-1999						
28	AM176762	177 bp mRNA	EST	16-NOV-1999						
29	AM176762	177 bp mRNA	EST	16-NOV-1999						
30	AM176762	177 bp mRNA	EST	16-NOV-1999						
31	AM176762	177 bp mRNA	EST	16-NOV-1999						
32	AM176762	177 bp mRNA	EST	16-NOV-1999						
33	AM176762	177 bp mRNA	EST	16-NOV-1999						
34	AM176762	177 bp mRNA	EST	16-NOV-1999						
35	AM176762	177 bp mRNA	EST	16-NOV-1999						
36	AM176762	177 bp mRNA	EST	16-NOV-1999						
37	AM176762	177 bp mRNA	EST	16-NOV-1999						
38	AM176762	177 bp mRNA	EST	16-NOV-1999						
39	AM176762	177 bp mRNA	EST	16-NOV-1999						
40	AM176762	177 bp mRNA	EST	16-NOV-1999						
41	AM176762	177 bp mRNA	EST	16-NOV-1999						
42	AM176762	177 bp mRNA	EST	16-NOV-1999						
43	AM176762	177 bp mRNA	EST	16-NOV-1999						
44	AM176762	177 bp mRNA	EST	16-NOV-1999						
45	AM176762	177 bp mRNA	EST	16-NOV-1999						

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 177)  
HCGP h<sup>+</sup> p://www.ludwig.org.br/ORESTES.  
The FAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
On May 18, 1998 this sequence version replaced gi:3138231.

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:15 : Search time 383.18 Seconds  
(without alignments)  
13.059 Million cell updates/sec

Title: US-08-956-991-6

Perfect score: 20  
Sequence: 1 cctgtatgacctgcaggaag 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	V31984	Human Down syndrome
2	20	100.0	388	V31982	Human Down syndrome
3	20	100.0	6604	V31981	Human Down syndrome
4	18.4	92.0	6413	V31988	Human Down syndrome
5	17.4	87.0	2173	V31987	Mouse Down syndrome
6	16.4	82.0	847	V59703	Human secreted pro
7	16.4	82.0	852	V59704	Human secreted pro
8	16.4	82.0	1250	X24827	Calcium activated
9	16.4	82.0	1287	V35463	Human hIK1 coding
10	16.4	82.0	1354	V59705	Human secreted pro
11	16.4	82.0	1486	V59523	Human secreted pro
12	16.4	82.0	1534	X24828	Calcium activated
13	16.4	82.0	2238	X24826	Calcium activated
14	16.4	82.0	2261	X24825	Human calcium acti
15	15.8	79.0	1285	Q20229	Sequence of the mu
16	15.8	79.0	1296	Q20228	ADP- glucose- pyroph
17	15.8	79.0	1296	Q80082	ADP- glucose- pyroph
18	15.8	79.0	1296	Q80083	ADP- glucose- pyroph
19	15.8	79.0	1332	Q25979	gluc. glycogen bio
20	15.8	79.0	1332	V23850	E. coli ADP- glucos
21	15.8	79.0	3955	T01590	Rat neu promoter.
22	15.4	77.0	24	Q45927	GCNA-2, GCNA bindi
23	15.4	77.0	24	Q45929	GCNA-4, GCNA bindi
24	15.4	77.0	5102	V10261	Human COL4A6 gene.
25	15.2	76.0	22	Q67130	CD40 ligand gene m
26	15.2	76.0	50	X52145	Synthetic plasmid
27	15.2	76.0	96	X03315	Test RNA fragment
28	15.2	76.0	717	V69319	Neomycin resistanc
29	15.2	76.0	818	V13141	np11 coding regio
30	15.2	76.0	1003	V13140	np11 CDS fused to
31	15.2	76.0	1186	Q25707	Chimeric neo gene
32	15.2	76.0	1321	V44078	mu-NDO-r DNA fragm
33	15.2	76.0	1415	Q28574	Sequence of the do
34	15.2	76.0	1521	V14348	Plasmid pCMVgfo11

35	15.2	76.0	1521	V14353	Plasmid pPKGfo25
36	15.2	76.0	1851	T90372	Recombinant molecu
37	15.2	76.0	1872	Q06828	Extracellular port
38	15.2	76.0	2118	T90371	Recombinant molecu
39	15.2	76.0	2224	T90373	Recombinant molecu
40	15.2	76.0	2239	X08974	relA-bacil encodin
41	15.2	76.0	2385	T72725	Her2-GM-CSF immuno
42	15.2	76.0	2447	Q30000	Chimeric lux::npt
43	15.2	76.0	2526	T62395	Construct pMK10/AS
44	15.2	76.0	2529	Q50489	Promoterless neo/a
45	15.2	76.0	2549	Q56107	DNA encoding a 11p

## ALIGNMENTS

RESULT 1	
V31984	
ID V31984 standard; cDNA; 20 BP.	
AC V31984:	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM primer B9-131R.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; PCR; primer; ss.	
OS Synthetic.	
OS Homo sapiens.	
PN WO9817795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; U19547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI Kolenberg, JR.	
DR WPI; 98-271791/24.	
PT New isolated Down's syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Example 5: Page 79: 109pp: English.	
CC Reverse primer B9-131R and forward primer B9-131F (see V31983)	
CC were used in RT-PCR assays of cDNA libraries of various human	
CC tissues to determine human Down syndrome-cell adhesion molecule	
CC DS-CAM (see V31981 and V31988) expression. The results	
CC demonstrated expression of human DS-CAM mRNA in foetal and adult	
CC brain, and foetal kidney. In addition, a breast carcinoma cell	
CC line showed expression of human DS-CAM mRNA. DS-CAM polypeptides	
CC (see W42086-87) are associated with developmental and neurological	
CC processes. The polypeptides and nucleic acids are used to develop	
CC products for the detection, diagnosis and therapy of developmental	
CC and neurological abnormalities.	
SQ Sequence 20 BP: 5 A; 5 C; 6 G; 4 T;	
Query Match 100.0%; Score 20; DB 1; Length 20;	
Best Local Similarity 100.0%; Pred. No. 0.77;	
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 cctgtatgacctgcaggaag 20	
DB 1 CCIGTATGACCTGCAGGAAG 20	
RESULT 2	
V31982/c	
ID V31982 standard; cDNA; 388 BP.	
AC V31982:	
DT 28-SEP-1998 (first entry)	
DE Human Down syndrome-cell adhesion molecule DS-CAM probe E51.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; human; ds; ss.	
OS Homo sapiens.	

PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.  
 PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities  
 PS Example 2: Page 79: 109pp: English.  
 CC This polynucleotide comprises cDNA fragment E51 that was isolated from a human trisomy 21 foetal brain (14-wk gestation) cDNA library following a modified direct cDNA selection technique applied to bacterial and pl artificial chromosomes between ER52 and MX1. E51 was used as a probe to screen the trisomy 21 foetal brain library. 62 clones were isolated from the 2 million clones of the original library. Overlapping clones were obtained that were used to deduce a full-length sequence (see V31981) coding for novel Down syndrome-cell adhesion molecule DS-CAM1 (see M42086). A splice variant, DS-CAM2 (see V31919), was also identified. DS-CAM polypeptides are associated with developmental and neurological processes. The polypeptides and nucleic acids are used to develop products for the detection, diagnosis and therapy of developmental and neurological abnormalities.  
 SO Sequence 388 BP: 83 A: 119 C: 95 G: 91 T:

Query Match 100.0%; Score 20; DB 1; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 0.94;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctgtatgacctgcaggaag 20  
 Db 331 CCTGTATGACCTGCAGGAG 312

RESULT 3  
 V31981  
 ID V31981 standard: cDNA: 6604 BP.  
 AC V31981;  
 DT 28-SEP-1998 (first entry)  
 DE Human Down syndrome-cell adhesion molecule DS-CAM1 cDNA.  
 KM DS-CAM1: Down syndrome-cell adhesion molecule; neural cell;  
 KM signal transduction; trisomy 21; mental retardation;  
 KM holoprosencephaly; corpus callosum agenesis;  
 KM schizencephaly; diagnosis: assay; human; ds; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 453..6185  
 FT sig\_peptide 452..521  
 FT mat\_peptide 522..6162  
 FT /\*tag= b  
 FT /\*tag= c  
 PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.  
 PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 DR P-PSDB: M42086.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities  
 PS Claim 3: Page 65-72: 109pp: English.  
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule DS-CAM1 (see M42086), a cell surface glycoprotein belonging to a novel subclass of the Ig superfamily with highest homology to neural cell adhesion molecules. A modified direct cDNA selection technique was applied to bacterial and pl artificial chromosomes between ER52 and MX1 by using cDNA from trisomy 21 human foetal

CC brain. A unique cDNA fragment, designated E51 (see V31982), was obtained and used to screen a trisomy 21 human foetal brain (14-wk gestation) cDNA library. Further clones were obtained by exon trapping. The complete DS-CAM1 cDNA sequence was deduced from overlapping clones. A splice variant cDNA (see V31988), encoding non-membrane bound DS-CAM2 (see M42087), was also identified. The CC DS-CAM gene spans 900-1200 kb of genomic DNA and maps at chromosome 21q22.2-22.3. The invention also provides murine DS-CAM partial sequences (see V31985-87), expression vectors and host cells, transgenic animals, antisense oligonucleotides, and primers useful CC for amplification of DS-CAM nucleic acid. DS-CAM polypeptides are CC associated with developmental and neurological processes. They can CC be used in e.g. neural prosthetic devices used in entubulation CC methods of repairing (regenerating) damaged or severed peripheral CC nerves. The products can also be used in detection, diagnosis and CC therapy of developmental and neurological abnormalities such as CC Down syndrome, mental retardation, holoprosencephaly, agenesis of CC the corpus callosum, or schizencephaly. Antisense oligonucleotides CC are used for inhibiting translation of mRNA.  
 SO Sequence 6604 BP: 1673 A: 1827 C: 1768 G: 1336 T:

Query Match 100.0%; Score 20; DB 1; Length 6604;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctgtatgacctgcaggaag 20  
 Db 5054 CCTGTATGACCTGCAGGAG 5073

RESULT 4  
 V31988  
 ID V31988 standard: cDNA: 6413 BP.  
 AC V31988;  
 DT 28-SEP-1998 (first entry)  
 DE Human Down syndrome-cell adhesion molecule DS-CAM2 cDNA.  
 KM DS-CAM2: Down syndrome-cell adhesion molecule; neural cell;  
 KM signal transduction; trisomy 21; mental retardation;  
 KM holoprosencephaly; corpus callosum agenesis;  
 KM schizencephaly; diagnosis: assay; human; ds; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 453..5168  
 FT /\*tag= a  
 PN W09817795-A1.  
 PD 30-APR-1998.  
 PF 23-OCT-1997: U19547.  
 PR 25-OCT-1996: US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 DR P-PSDB: M42087.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to develop products for detection, diagnosis and therapy of developmental and neurological abnormalities  
 PS Claim 3: Page 83-90: 109pp: English.  
 CC This cDNA clone codes for Down syndrome-cell adhesion molecule DS-CAM2 (see M42087), an extracellular, soluble protein belonging to a novel subclass of the Ig superfamily with highest homology to neural cell adhesion molecules. A modified direct cDNA selection technique was applied to bacterial and pl artificial chromosomes between ER52 and MX1 by using cDNA from trisomy 21 human foetal brain. A unique cDNA fragment, designated E51 (see V31982), was obtained and used to screen a trisomy 21 human foetal brain (14-wk gestation) cDNA library. Further clones were obtained by exon trapping, and the complete DS-CAM2 cDNA sequence was identified. DS-CAM2 is a splice variant of membrane-bound DS-CAM1 (see M42086) and is deleted for the entire transmembrane domain. The DS-CAM CC gene spans 900-1200 kb of genomic DNA and maps at chromosome 21q22.2-22.3. The invention also provides murine DS-CAM partial CC sequences (see V31985-87), expression vectors and host cells, transgenic animals, antisense oligonucleotides, and primers useful

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:22 ; Search time 216.06 Seconds  
(Without alignments)  
11.082, 111100 cell updates/sec

Title: US-08-956-991-6

Perfect score: 20  
Sequence: 1 cctgtatgacctgcaggaag 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9.COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	1296	1 US-08-090-523-1	Sequence 1, Appl1
2	15.8	79.0	1296	1 US-08-090-523-3	Sequence 3, Appl1
3	15.8	79.0	1296	1 US-08-398-627-1	Sequence 1, Appl1
4	15.8	79.0	1296	1 US-08-398-627-3	Sequence 3, Appl1
5	15.8	79.0	1296	1 US-08-406-858-1	Sequence 1, Appl1
6	15.8	79.0	1296	1 US-08-406-858-3	Sequence 3, Appl1
7	15.8	79.0	1296	6 PCT-US91-04036-1	Sequence 1, Appl1
8	15.8	79.0	1296	6 PCT-US91-04036-3	Sequence 3, Appl1
9	15.8	79.0	1296	6 PCT-US94-05275-1	Sequence 1, Appl1
10	15.8	79.0	1296	6 PCT-US94-05275-3	Sequence 3, Appl1
11	15.8	79.0	1323	1 US-07-735-065-3	Sequence 3, Appl1
12	15.8	79.0	1323	2 US-08-469-202-13	Sequence 13, Appl1
13	15.8	79.0	1323	4 US-08-484-434C-13	Sequence 13, Appl1
14	15.8	79.0	3955	1 US-08-229-515A-14	Sequence 14, Appl1
15	15.8	79.0	3955	1 US-08-645-865-14	Sequence 1, Appl1
16	15.4	77.0	24	6 PCT-US93-00321-2	Sequence 2, Appl1
17	15.4	77.0	24	6 PCT-US93-00321-4	Sequence 4, Appl1
18	15.4	77.0	5102	2 US-08-494-168-1	Sequence 1, Appl1
19	15.2	76.0	22	1 US-08-184-422-15	Sequence 15, Appl1
20	15.2	76.0	1186	1 US-08-064-121-2	Sequence 2, Appl1
21	15.2	76.0	1186	2 US-08-478-015-2	Sequence 2, Appl1
22	15.2	76.0	1186	2 US-08-475-975-2	Sequence 2, Appl1
23	15.2	76.0	1415	1 US-08-257-341-8	Sequence 8, Appl1
24	15.2	76.0	1853	1 US-08-404-732A-6	Sequence 6, Appl1
25	15.2	76.0	2119	5 US-09-032-372-6	Sequence 6, Appl1
26	15.2	76.0	2120	1 US-08-404-732A-4	Sequence 4, Appl1
27	15.2	76.0	2220	3 US-08-864-224-1	Sequence 1, Appl1

28	15.2	76.0	2224	1 US-08-404-732A-8	Sequence 8, Appl1
29	15.2	76.0	2285	4 US-09-146-283-3	Sequence 3, Appl1
30	15.2	76.0	2445	1 US-08-122-520C-8	Sequence 8, Appl1
31	15.2	76.0	3176	7 5212080-1	Patent No. 5212080
32	15.2	76.0	3200	1 US-08-453-104-23	Sequence 23, Appl1
33	15.2	76.0	3200	3 US-08-694-824-23	Sequence 23, Appl1
34	15.2	76.0	3201	1 US-08-453-104-22	Sequence 22, Appl1
35	15.2	76.0	3201	3 US-08-694-824-22	Sequence 22, Appl1
36	15.2	76.0	3740	1 US-07-794-400-14	Sequence 14, Appl1
37	15.2	76.0	3740	1 US-08-041-648-3	Sequence 3, Appl1
38	15.2	76.0	3740	1 US-08-217-529-3	Sequence 3, Appl1
39	15.2	76.0	3740	1 US-08-397-470-14	Sequence 14, Appl1
40	15.2	76.0	3768	3 US-08-625-101-1	Sequence 1, Appl1
41	15.2	76.0	3768	3 US-08-356-786-1	Sequence 1, Appl1
42	15.2	76.0	3974	5 US-09-026-343-33	Sequence 33, Appl1
43	15.2	76.0	4473	4 US-09-048-804-1	Sequence 1, Appl1
44	15.2	76.0	4530	1 US-08-229-515A-9	Sequence 9, Appl1
45	15.2	76.0	4530	1 US-08-645-865-9	Sequence 9, Appl1

## ALIGNMENTS

RESULT 1  
US-08-090-523-1  
; Sequence 1, Application US/06090523  
; Patent No. 5498830  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Stark, David M.  
; TITLE OF INVENTION: Enhanced Starch Biosynthesis  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Grace L. Bonner, Monsanto Co. B94F  
; STREET: 700 Chesterfield Parkway No. 5498830th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/090,523  
; FILING DATE: 19930712  
; CLASSIFICATION: 800  
; PRIOR APPLICATION NUMBER: US 07/709663  
; FILING DATE: 07-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/539763  
; FILING DATE: 18-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bonner, Grace L.  
; REGISTRATION NUMBER: 32,963  
; REFERENCE/DOCKET NUMBER: 38-21(10559)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314) 537-7286  
; TELEFAX: (314) 537-6047  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1296 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1293  
; US-08-090-523-1

Query Match 79.0%: Score 15.8; DB 1: Length 1296;  
Best Local Similarity 89.5%: Pred. No. 27;  
Matches 17: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ctgtatgacctgcaggaag 20  
||||| ||| ||| |||  
DB 667 CTGTATGAAGCTGTGAAG 685

## RESULT 2

US-08-090-523-3  
Sequence 3, Application US/08090523  
Patent No. 5498830  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Enhanced Starch Biosynthesis  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grace L. Bonner, Monsanto Co. B54F  
STREET: 700 Chesterfield Parkway No. 5498830th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/090.523  
FILING DATE: 19930712  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/709663  
FILING DATE: 07-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/539763  
FILING DATE: 18-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Bonner, Grace L.  
REGISTRATION NUMBER: 32,963  
REFERENCE/DOCKET NUMBER: 38-21(10559)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314) 537-7286  
TELEFAX: (314) 537-6047  
INFORMATION FOR SEQ. ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1293  
US-08-090-523-3

Query Match 79.0%: Score 15.8; DB 1: Length 1296;  
Best Local Similarity 89.5%: Pred. No. 27;  
Matches 17: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ctgtatgacctgcaggaag 20  
||||| ||| ||| |||  
DB 667 CTGTATGAAGCTGTGAAG 685

RESULT 3

US-08-398-627-1  
Sequence 1, Application US/08398627  
Patent No. 5608149  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Enhanced Starch Biosynthesis  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grace L. Bonner, Monsanto Co. B54F  
STREET: 700 Chesterfield Parkway No. 5608149th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/398.627  
FILING DATE: 03-MAR-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/090.523  
FILING DATE: 12-JUL-1993  
APPLICATION NUMBER: US 07/709663  
FILING DATE: 07-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/539763  
FILING DATE: 18-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Bonner, Grace L.  
REGISTRATION NUMBER: 32,963  
REFERENCE/DOCKET NUMBER: 38-21(10559)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314) 537-7286  
TELEFAX: (314) 537-6047  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1293  
US-08-398-627-1

Query Match 79.0%: Score 15.8; DB 1: Length 1296;  
Best Local Similarity 89.5%: Pred. No. 27;  
Matches 17: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ctgtatgacctgcaggaag 20  
||||| ||| ||| |||  
DB 667 CTGTATGAAGCTGTGAAG 685

## RESULT 4

US-08-398-627-3  
Sequence 3, Application US/08398627  
Patent No. 5608149  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Enhanced Starch Biosynthesis  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:40 ; Search time 4987.54 Seconds  
(without alignments)  
15.140 Million cell updates/sec

Title: US-08-956-991-6  
Perfect score: 20  
Sequence: 1 ccgtatgacctgcaggaag 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues  
Total number of hits satisfying chosen parameters: 5077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: qb\_est1:\*  
21: qb\_est2:\*  
22: qb\_est3:\*  
23: qb\_est4:\*  
24: qb\_est5:\*  
25: qb\_est6:\*  
26: qb\_est7:\*  
27: qb\_est8:\*  
28: qb\_est9:\*  
29: qb\_est10:\*  
30: qb\_est11:\*  
31: qb\_est12:\*  
32: qb\_est13:\*  
33: qb\_est14:\*  
34: qb\_est15:\*  
35: qb\_est16:\*  
36: qb\_est17:\*  
37: qb\_est18:\*  
38: qb\_est19:\*  
39: qb\_est20:\*  
40: qb\_est21:\*  
41: qb\_est22:\*  
42: qb\_est23:\*  
43: qb\_est24:\*  
44: qb\_est25:\*

45: qb\_est26:\*  
46: qb\_est27:\*  
47: qb\_est28:\*  
48: qb\_est29:\*  
49: qb\_est30:\*  
50: qb\_est31:\*  
51: qb\_est32:\*  
52: em\_est10:\*  
53: em\_est11:\*  
54: em\_est12:\*  
55: em\_est13:\*  
56: em\_est14:\*  
57: em\_est15:\*  
58: em\_est16:\*  
59: qb\_est13:\*  
60: qb\_est14:\*  
61: qb\_est15:\*  
62: qb\_est16:\*  
63: qb\_est17:\*  
64: qb\_est18:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: qb\_est19:\*  
70: qb\_est140:\*  
71: qb\_est141:\*  
72: qb\_est142:\*  
73: qb\_est143:\*  
74: qb\_est144:\*  
75: em\_est131:\*  
76: em\_est132:\*  
77: em\_est133:\*  
78: em\_est134:\*  
79: qb\_gss1:\*  
80: qb\_gss2:\*  
81: qb\_gss3:\*  
82: qb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: qb\_gss5:\*  
88: qb\_gss6:\*  
89: qb\_gss7:\*  
90: qb\_gss8:\*  
91: qb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: qb\_gss10:\*  
100: qb\_gss11:\*  
101: em\_gss12:\*  
102: qb\_gss12:\*  
103: qb\_gss13:\*  
104: qb\_gss14:\*  
105: qb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	188	50	F25472
2	16.8	84.0	334	61	A1841727

F25472 HSPD12539 H  
A1841727 UI-M-AL0-

```

C 3 16.8 84.0 338 21 F00929
4 16.8 84.0 339 39 AA855923
C 5 16.8 84.0 349 21 R05334
6 16.8 84.0 349 61 A1836340
C 7 16.8 84.0 350 62 A1905927
8 16.8 84.0 416 35 AA656681
9 16.8 84.0 485 87 AO815436
10 16.8 84.0 516 34 AA498668
11 16.8 84.0 522 30 AA204382
12 16.8 84.0 550 42 A1116267
13 16.8 84.0 571 103 A0438215
C 14 16.8 84.0 593 63 A1988028
15 16.8 84.0 600 62 A1905935
16 16.8 84.0 649 74 AA177833
17 16.8 84.0 707 35 AA545906
18 16.8 84.0 734 60 A1788686
19 16.8 84.0 745 40 AA980418
C 20 16.8 84.0 788 60 A1788722
21 16.4 82.0 253 28 AA066808
22 16.4 82.0 260 90 AA058631
23 16.4 82.0 410 28 AA076337
24 16.4 82.0 479 74 AA201060
C 25 16.4 82.0 509 74 AA200065
26 16.4 82.0 515 105 A0600786
C 27 16.4 82.0 704 59 A1768257
28 16.4 82.0 713 88 A0853594
C 29 16.4 82.0 738 82 A0889443
30 16.4 80.0 575 63 A1997586
C 31 15.8 79.0 133 26 W55978
32 15.8 79.0 166 30 AA247199
33 15.8 79.0 180 22 R58397
34 15.8 79.0 186 22 R58397
35 15.8 79.0 200 59 A1769721
C 36 15.8 79.0 217 25 W10919
37 15.8 79.0 235 25 N87427
38 15.8 79.0 242 22 R57457
39 15.8 79.0 248 30 AA216055
40 15.8 79.0 255 25 N84658
41 15.8 79.0 257 31 AA328566
42 15.8 79.0 259 25 N85521
43 15.8 79.0 261 33 AA382414
44 15.8 79.0 267 29 AA196406
45 15.8 79.0 272 27 C05020

```

## ALIGNMENTS

```

RESULT 1
F25472 188 bp mRNA EST 13-MAY-1999
DEFINITION HSD12539 HM3 Homo sapiens cDNA clone s4000038E11, mRNA sequence.
VERSION F25472.1 GI:4811098
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 188)
AUTHORS Lanfranchi,G., Murao,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
Pandofo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
TITLE Identification of 4370 expressed sequence tags from a
3'-end-specific cDNA library of human skeletal muscle by DNA
sequencing and filter hybridization
JOURNAL Genome Res. 6 (1), 35-42 (1996)
MEDLINE 96276048
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246745.
CONTACT: Valle G.
CRIBI Biotechnology Centre
University of Padua
Via Trieste 75, 35121 Padua, Italy
ABI Chromatograms and other information are available on WWW at

```

http://grup.bio.unipd.it.  
Location/Qualifiers

## FEATURES

## source

```

1..188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="s4000038E11"
/clone_lib="HM3"
/sex="female"
/tissue-type="pectoral muscle (after mastectomy)"
/note="Vector: pCDNAII (Invitrogen); Site_1: BstXI;
Site_2: NotI. The library is not subtracted nor normalized.
Lanfranchi. This library is not subtracted nor normalized.
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer
(5'-biotin-AACCGCGCTCGAGCGCGCTTTTCTTTTCTTTT-3'). The
ds cDNA was sonicated and size-selected in the range
350-550 bp. The 3' specific fragments were selected by
streptavidin coated magnetic beads, ligated to
non-palindromic BstXI adapters, NotI digested and
directionally cloned into BstXI-NotI cut pCDNAII vector."
BASE COUNT 67 a 44 c 49 g 28 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 50; Length 188;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 cctgtatgacctgcaggaag 20
Db 73 cctgtatgacctgcaggaag 92

```

```

RESULT 2
A1841727 334 bp mRNA EST 14-JUL-1999
DEFINITION UT-M-A10-abo-f-04-0-UI s1 NIH-BMAP_MCO Mus musculus cDNA clone
UT-M-A10-abo-f-04-0-UI 3', mRNA sequence.
VERSION A1841727
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 334)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On May 9, 1996 this sequence version replaced gi:1132832.
CONTACT: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized prefrontal cortex library cDNA library preparation:
M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be
made available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA-Tes.

```

FEATURES  
source Location/Qualifiers  
1..334

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:38:52 ; Search time 7810.03 seconds  
(without alignments)  
-327.353 Million cell updates/sec

Title: US-08-956-991-7  
Perfect score: 842  
Sequence: 1 ccggcgccggcgccggcgag.....cgcgtgagcgtccttcgca 842

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da1: \*  
2: gb\_da2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pr1: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_vi: \*  
17: em\_fun: \*  
18: em\_hum1: \*  
19: em\_hum2: \*  
20: em\_in: \*  
21: em\_om: \*  
22: em\_or: \*  
23: em\_ov: \*  
24: em\_pat: \*  
25: em\_ph: \*  
26: em\_pl: \*  
27: em\_ro: \*  
28: em\_sts: \*  
29: em\_sy: \*  
30: em\_un: \*  
31: em\_vi: \*  
32: gb\_htg1: \*  
33: gb\_htg2: \*  
34: gb\_in1: \*  
35: gb\_in2: \*  
36: em\_bal: \*  
37: em\_da2: \*  
38: em\_hum3: \*  
39: em\_hum4: \*  
40: gb\_pr4: \*  
41: gb\_htg3: \*  
42: gb\_htg4: \*  
43: gb\_htg5: \*  
44: gb\_htg6: \*

45: gb\_htg7: \*  
46: em\_htg1: \*  
47: em\_htg2: \*  
48: em\_htg3: \*  
49: em\_hum5: \*  
50: gb\_pl3: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	670.2	79.6	6413	11	AF023450 Homo sapi
2	594	70.5	6110	11	AF023449 Homo sapi
3	292.2	34.7	119678	11	AF064866 Homo sapi
4	292.2	34.7	120007	11	AF064864 Homo sapi
5	148.2	17.6	189607	32	AP000757 Homo sapi
6	148.2	17.6	223724	32	AP000834 Homo sapi
7	125.6	14.9	155407	11	AF042090 Homo sapi
8	77.6	9.2	109866	11	AF043945 Homo sapi
9	58.8	7.0	124395	44	AC013397 Homo sapi
10	51.4	6.1	66362	43	AC013815 Homo sapi
11	49	5.8	79333	35	AC005452 Homo sapi
12	46.2	5.5	69128	45	AC017764 Homo sapi
13	41.8	5.0	107475	45	AC017106 Homo sapi
14	41	4.9	1970	4	GGY17794 Gallus gall
15	41	4.9	162648	43	AC010827 Homo sapi
16	40.4	4.8	417	34	AB017336 Homo sapi
17	40	4.8	80908	11	HS524E15 Homo sapi
18	39.8	4.7	308292	45	AC012384 Homo sapi
19	39.4	4.7	73511	11	HS65A6 Homo sapi
20	39.4	4.7	1237	9	AP000353 Homo sapi
21	39.2	4.7	66	16	HEH5Y1G3 Homo sapi
22	39.2	4.7	7660	40	AF116458 Homo sapi
23	39.2	4.7	12001	5	AR048721 Homo sapi
24	39.2	4.7	26245	16	HS1US Homo sapi
25	39.2	4.7	26245	16	HS1US Homo sapi
26	39.2	4.7	43034	11	HSN5H6 Homo sapi
27	39.2	4.7	152261	16	HE1C Homo sapi
28	39.2	4.7	152261	16	HE1C Homo sapi
29	38.6	4.6	5887	9	SS1132828 Homo sapi
30	38.6	4.6	177241	11	HS402G11 Homo sapi
31	38.4	4.6	1959	5	IO1837 Homo sapi
32	38.4	4.6	3573	5	SC1NHBAR Homo sapi
33	38.4	4.6	3588	5	IO52233 Homo sapi
34	38.4	4.6	3588	5	IO52233 Homo sapi
35	38.4	4.6	3588	5	IO52233 Homo sapi
36	38.4	4.6	3588	5	IO52233 Homo sapi
37	38.4	4.6	3588	5	IO52233 Homo sapi
38	38.2	4.5	1525	4	CHKESTRP13 Homo sapi
39	38.2	4.5	7618	4	XB28SRNA Homo sapi
40	38.2	4.5	81463	45	AC016882 Homo sapi
41	38	4.5	1478	1	SC110F3 Homo sapi
42	38	4.5	1524	5	AR065197 Homo sapi
43	38	4.5	2900	1	PG1IP Homo sapi
44	38	4.5	2900	5	AL6323 Homo sapi
45	38	4.5	2900	5	AR038049 Homo sapi

#### ALIGNMENTS

RESULT 1  
AF023450 AF023450 6413 bp mRNA  
LOCUS Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (USCAM)  
DEFINITION mRNA, complete cds.  
ACCESSION AF023450  
VERSION AF023450.1 GI:3169767













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:44 ; Search time 4987.54 Seconds  
(without alignments)  
637.410 Million cell updates/sec

Title: US-08-956-991-7

Perfect score: 842  
Sequence: 1 ccggcgccggcgccggcgga.....cgctggagctgcttgcga 842

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est33:\*  
53: gb\_est34:\*  
54: gb\_est35:\*  
55: gb\_est36:\*  
56: gb\_est37:\*  
57: gb\_est38:\*  
58: gb\_est39:\*  
59: gb\_est40:\*  
60: gb\_est41:\*  
61: gb\_est42:\*  
62: gb\_est43:\*  
63: gb\_est44:\*  
64: gb\_est45:\*  
65: gb\_est46:\*  
66: gb\_est47:\*  
67: gb\_est48:\*  
68: gb\_est49:\*  
69: gb\_est50:\*  
70: gb\_est51:\*  
71: gb\_est52:\*  
72: gb\_est53:\*  
73: gb\_est54:\*  
74: gb\_est55:\*  
75: gb\_est56:\*  
76: gb\_est57:\*  
77: gb\_est58:\*  
78: gb\_est59:\*  
79: gb\_est60:\*  
80: gb\_est61:\*  
81: gb\_est62:\*  
82: gb\_est63:\*  
83: gb\_est64:\*  
84: gb\_est65:\*  
85: gb\_est66:\*  
86: gb\_est67:\*  
87: gb\_est68:\*  
88: gb\_est69:\*  
89: gb\_est70:\*  
90: gb\_est71:\*  
91: gb\_est72:\*  
92: gb\_est73:\*  
93: gb\_est74:\*  
94: gb\_est75:\*  
95: gb\_est76:\*  
96: gb\_est77:\*  
97: gb\_est78:\*  
98: gb\_est79:\*  
99: gb\_est80:\*  
100: gb\_est81:\*  
101: gb\_est82:\*  
102: gb\_est83:\*  
103: gb\_est84:\*  
104: gb\_est85:\*  
105: gb\_est86:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177.2	21.0	619	79	FR0002941	286724 F.rubripes
2	135	16.0	147	79	HSWC18B12	X88325 H.sapiens D

source	1. .519	21.0%	Score 177.2;	DB 79;	Length 619;	Best Local Similarity 75.6%;	Pred. No. 6.7e-36;	Mismatches 73;	Indels 1;	Gaps 1;
BASE COUNT	133 a	169 c	141 g	156 t	20 others					
ORIGIN										
AI498283	tm85h11.x									
AL027078	Fugu rubripes									
AL053013	Drosophila									
AL098882	Drosophila									
AL065829	Drosophila									
AL054280	Drosophila									
AL135788	qvl3b02.x									
AL100255	Drosophila									
AL085798	qbe0011F									
AL066051	Drosophila									
AL071731	Drosophila									
AL0327293	nbx0040P									
AL065254	Drosophila									
AL109149	Drosophila									
AL103118	Drosophila									
AL129175	qhl6804.x									
AL133984	q33b02.x									
AL103945	Drosophila									
AL098749	Drosophila									
AL135609	qy64f03.x									
AL106535	Drosophila									
AL099352	Drosophila									
AL071004	Drosophila									
AL108419	Drosophila									
AL106580	Drosophila									
D60890	Hum13600B									
AL165949	tt41g08.x									
AG866010	nbeb0026L									
AG894372	HS_4832.A									
AL108460	Drosophila									
AL108775	Drosophila									
AG896580	HS_5440.A									
AL059740	Drosophila									
AL053409	Drosophila									
AL066742	Drosophila									
AL077161	Drosophila									
AL072149	Drosophila									
AI498283	47	AI498283	13.2	450						
FR0030709	79	FR0030709	7.8	522						
CNS0091P	79	CNS0091P	5.2	925						
CNS010EW	79	CNS010EW	5.2	1009						
CNS0060N	79	CNS0060N	4.3	910						
CNS004NB	79	CNS004NB	4.2	839						
AI135788	79	AI135788	4.1	641						
CNS011H1	45	CNS011H1	4.9	788						
AO857989	79	AO857989	4.9	794						
CNS006XK	79	CNS006XK	4.9	935						
W42199	26	W42199	4.8	1049						
AI107879	42	AI107879	4.7	587						
CNS0091P	79	CNS0091P	4.7	925						
CNS016P9	79	CNS016P9	4.7	1064						
CNS00G7F	79	CNS00G7F	4.7	1101						
Q327293	100	Q327293	4.6	764						
CNS003WG	79	CNS003WG	4.6	964						
CNS018C3	79	CNS018C3	4.6	237						
CNS0130K	79	CNS0130K	4.6	472						
AI1249175	44	AI1249175	4.6	413						

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:47:15 : Search time 7810.03 Seconds  
(without alignments)  
-349.125 Million cell updates/sec

Title: US-08-956-991-8

Sequence: 898 1 tgcgcgcgcgttcgaagcct.....gcgccacacgttcgtcta 898

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_om:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	502	55.9	6110	11 AF023449	AF023449 Homo sapi
C 2	502	55.9	6413	11 AF023450	AF023450 Homo sapi
C 3	244.8	27.3	6025	10 AB032958	AB032958 Homo sapi
C 4	170	18.9	159424	11 AF064865	AF064865 Homo sapi
C 5	74	8.2	193789	32 AP000849	AP000849 Homo sapi
C 6	39	4.3	1648	34 PIRPPI	X01469 Plasmidium
C 7	38.8	4.3	1238	5 AR064010	AR064010 Sequence
C 8	38.8	4.3	1271	11 HS057099	US7099 Human APB-
C 9	38.8	4.3	2793	5 AR064011	AR064011 Sequence
C 10	38	4.2	96870	41 AC009739	AC009739 Drosophila
C 11	37.8	4.2	10915	2 AE001900	AE001900 Deinococc
C 12	37.8	4.2	103445	32 AP000644	AP000644 Homo sapi
C 13	37.6	4.2	855	34 PFRHCR	M15317 P. lophurae
C 14	37.6	4.2	1472	3 HRSIGMR	L49414 Equus caball
C 15	37.2	4.1	92797	11 AF038458	AF038458 Homo sapi
C 16	37.2	4.1	148432	35 AC004642	AC004642 Drosophila
C 17	36.6	4.1	408	1 UKBNIFD2	D26272 Unknown nit
C 18	36.4	4.1	1408	7 AB008609	U08609 Atropa bell
C 19	36.4	4.1	140	7 NSU08616	U08616 Nolana spat
C 20	36.4	4.1	721	5 AR06494	AR06494 Sequence
C 21	36.2	4.0	1308	5 AR064008	AR064008 Sequence
C 22	36.2	4.0	1308	12 RN057097	US7097 Rattus norv
C 23	35.8	4.0	3971	4 CHKCONNE	D16541 Chicken DNA
C 24	35.8	4.0	1345	7 STZCPBCL	L12658 Streptozia
C 25	35.2	3.9	1400	8 MGICPRBCL	X54345 M. macrophyll
C 26	35.2	3.9	1428	7 CYCPRBCL	L01900 Cyrtilla fac
C 27	35.2	3.9	1428	7 ENACBCL	L12616 Enkianthus
C 28	35.2	3.9	1428	7 MAICPRBCL	L12655 Magnolia hy
C 29	35.2	3.9	1428	7 MAICPRBCL	L12656 Magnolia sa
C 30	35.2	3.9	1428	7 MAICPRBCL	L12659 Michelia fi
C 31	35.2	3.9	1428	7 MAICPRBCL	L12666 Talium ova
C 32	35.2	3.9	1428	7 MAICPRBCL	Y00815 Human mRNA
C 33	35.2	3.9	1428	7 MAICPRBCL	AL079356 Streptomy
C 34	35.2	3.9	1428	7 MAICPRBCL	U92813 Hirudo med
C 35	35.2	3.9	1428	7 MAICPRBCL	AC012006 Homo sapi
C 36	35.2	3.9	1428	7 MAICPRBCL	AC004952 Homo sapi
C 37	35	3.9	137968	33 AC004952	AF123662 Zaluzians
C 38	35	3.9	137968	33 AC004952	AF123662 Zaluzians
C 39	34.8	3.9	1322	8 AF161798	AF161798 Orobanch
C 40	34.8	3.9	1389	8 AF161798	U08608 Anthocercis
C 41	34.8	3.9	1408	7 AVU08608	U08613 Lycium cest
C 42	34.8	3.9	1408	7 LCU08613	U08614 Mandragora
C 43	34.8	3.9	1408	7 MPO08615	U08615 Nicandra ph
C 44	34.8	3.9	1408	7 MPO08615	U08615 Nicandra ph
C 45	34.8	3.9	2001	34 DM025344	U25344 Drosophila

#### ALIGNMENTS

RESULT 1  
AF023449/c AF023449 6110 bp mRNA PRI 01-JUN-1998  
LOCUS Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)  
DEFINITION mRNA, partial cds.  
ACCESSION AF023449  
VERSION AF023449.1 GI:3169765

KEYWORD	source
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 6110) Yamakawa,K., Huo,Y.-K., Haendel,M.A., Hubert,R., Chen,X.-N., Lyons,G.E. and Korenberg,J.R. DSCAM: A Novel Member of the Immunoglobulin Superfamily Maps in a Down Syndrome Region and is Involved in the Development of the Nervous System
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 6110) Yamakawa,K., Huo,Y.-K., Haendel,M.A., Hubert,R., Chen,X.-N., Lyons,G.E. and Korenberg,J.R. Direct Submission Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research Institute, 110 George Burns Road, Davis Building, Suite 2005, Los Angeles, CA 90048-1869, USA
TITLE	Location/Qualifiers
FEATURES	1..6110
source	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21 (trisomy 21)" /map="21q22, between HMG14 and MX1" /tissue-type="brain" /dev_stage="14 weeks, fetal" /clone="CHD2-42" /note="derived from alternately-spliced mRNA" 1..6110 /gene="DSCAM" /gene="DSCAM" /gene="DSCAM" /note="member of immunoglobulin superfamily; involved in nervous system development" /codon_start=1 /product="Down syndrome cell adhesion molecule" /protein_id="AAC17966.1" /db_xref="GI:3169766" /translation="VFSEDLHSSLYFNASLQEVYFASTGTGLTPCPACIGPPTVTLRM LHAGEIYDVPGRHVAHPNCTLOIFPEPPSESTLHDNTVYCTAIPNSGKLRSDV HKLALREPIYVREDOKTNGVNAVFCKLIPSSVEAYIVVSEKDTVSLVSGSR ITSGCALIKVQVNDGGLYNNRCITRHHYVETGTSNSARLIVFADPNMSPSLIDGDF HRKMAAGRVLPCKALGSGRPDXYRMLKDNMPELRSRFOKTVGLILNIRPDSGS YVCEASNRTAKTVLGRLYVQPLKATISPRKYSVSGVSYVSLCVSTGEODELSMY RNGELINQKNNVRIITGINHEMLINDHMKPSDGAHQCFVRKDKLSADYVQVLEQCT PFIIAFSEKVVNPALPEVSLMNCNKGTPLPITVTLDDDDPLTKGSHRISQMITSEGN VYVYINSSQVGDGGRVRCITANNSAGVLAQVLAIRNGAPASIRPMKNTAIGADRTY IHCVRIQPYYSIKWYKNSNLLPENHRQVAFENNGLTLSDVQKVEDEGETCNVLYQ POLSTQSQVHVTYKRPPILOPEEPREFSIGORVLPQVVGSDLPITITWOKGRPLP GSLGVTIHDPTSSLRKISLNSLHMGNGVETIAINBAVAHQSOLIVARPKRVQOP RQDDIYGAVALNCSAGRPVPTTVYKFSKGAQVPOPIALNGRITLNSGLIK HVEDSDGYILCKVINDVDGADVSKSMYLTVAIPMIIYSIPVITLATLOGKKENSTAH GKPKILVWKEKEDRIINPEMARLYSTVEVEEVIYSTQLLPVRESGFFSCNALSIN YEDGRIIOLIVQEPDPPEIEIKDVKARTITLRWTGFDNSPIITGDIENKRSOSI WMSADRTKQVSPOLNSATIIDIDHPESTYSIMYKANRIGSEPNELTIADEAPARISG PQEVHILBPISSQIRVWMAKPKALONGIILRGVIGREYESTGYNFOFNILSYDVG DESEVITLDMLNKFTOYGLVQACNACAGPSSQSEITITLTDVPSYEPENQAATASP EESITLSWTLKSEALNGILQGRVITANMDDELGEIKNTTPOSLTEIGLEKYN YSIOVLAFTRAGDGVRSQEIFRTKEDVPGPPAGVKAASAVFYFSLPRLKLNIT IKRYVIEFSHPYPIVISEFASPDSEFRIAPLSNNROYSVAAVYISAGNSSELIIT TYELAKAPARILITFSGVTTPPMKMDIILPCAVADPSPAVKMKMSDNGSTLVIDG RSTISNGSFIILRTAKAGDSGYSCILANNMGSDEIILNLQVQPDQPLRLVTKTSS SITLISLMDNGSGSIRGILQVSEDSNGSPISIPSESRYLEMLKCGWYKFTTS LILPQNGVFGRTVITTEAKTLEPOFSKEDLPFASINTTTRRLNLIGNMDGGPITS FLERPGFGRVITWTAORTLSKSTILTDLOEATYELQKVCNAGCAEKAQNAFALL INDGSTIPLIKSVQVNEGLITNGELKLVITSLVGLVLEPVLVLVRRRRRDE LKLIDAKSLAELMSKNTRTSDTLSSKOOTLRMIDILPRQOLLIEEDTETLIDDR TYLLTADGGEAKOKSLVTHIVHVSQAQTCPLVYVSDASBPSTNTRRNKAG TANRRAQOMTLNRPRTISHTLITDMLPLPPRAAGSVDKSDSYVSPQDDIDRAK SSAVSTESASSTYEEIARAYEHAHKEOOLRHKKFTITTECFISITSEOLTAGTNEYT SLTSTSPSESGICORTFASPCKPDGGRVNMNAVPRALIQVYISYICLHLHMTFC"
BASE COUNT	1621 a 1658 g 1283 t

ORIGIN

Query Match	Similarity	Score	DB	Length
Best Local	82.3%	Pred. No. 8.2e-140		
Matches	749	Conservative	0	Mismatches 135; Indels 26; Gaps 14
3	ccgcgcggtgtgaacactgtacta--caggccatactgcgtga--ttacaggtgtcca	59		
Db	3169 GGGCCCGGTACAGGCGCTCCACACAGGCGCTGACTGAGTGAACCTATTACAGGTGTCCA	3110		
Qy	60 ggggtacacttcgctgt-cccggggtgtcaactactaataatgtatgttgaactgaaagttaac	118		
Db	3109 GGGGTAAACCCACACTGTCCCGCTGGGTGTGACACTGATTAATGTTGAATTGGAAGTTAC	3050		
Qy	119 ccc---gtctctaccccgtagacctatgtttaacgagaaatgatcccgctgtataag	174		
Db	3049 CCCCAGTGTGACTGACTGGTAACCTATTGTAAGCAGCAAGTAATCCATTGTCGAAT	2990		
Qy	175 tgtctcttggaagcctctccagatgaaccctgatactctgataagaagtgtgtccaaatga	234		
Db	2989 GATTCTTGGGAGCCT--TCGATGAGCCCTGATGCTCTGAGATGATTAAGCTTCAGAGTGA	2931		
Qy	235 actctctgaggtaga--caccacagactgctccatccgcgcgtatagtgtatcctgtctg	293		
Db	2930 ACTTCTGAGGTGAGACATCAGAGGCTGCTGCTCCGCGTGAAGTGAAGCTGTGCTG	2871		
Qy	294 ggtcacccttgccaatccggtctcttggtgaatgtcgatgtgtfaggtgtgaggaagag	353		
Db	2870 GCGTCGCTCTTGCCAAATCCCGTTCCTGGCGCTACATCGGATGCTGTAGAGGAGGAAGG	2811		
Qy	354 tggatcatcaatgtatgtgtgcccaggtcagcttgaggggaaacatcttgytctctaga	413		
Db	2810 TGGATATCATGATGCTGGCGGCACTTCACTGAGGGGAACATCTTGGTCTCTGACGA	2751		
Qy	414 gaatcccaacagtcgtattatcttgacatccaactgtatagctgtgaaatgagggctgt	473		
Db	2750 GAATCCCGAGGAGTGTGAATTAATTTTTCAT--CAATATGTAGCCGTGATAGGAGATGT	2693		
Qy	474 tggcatcaaaccccatggtctcacactgagctgtatgtgtgagactttacactctctgac	533		
Db	2692 TTGATATTAACCCCATGTGCTCACCTGAGCTAATTTGCGTGCTTTGACATCT--TTGATC	2634		
Qy	534 tcaatctccggagatctggtgtgttcttgcaatgtgagttgaattatctccacggtctcc	593		
Db	2693 TCAATTTCCGGAGCGCTGGCGGCTCTTGACACTGTGAGCTGAATTAATTCACAGGCTCTCC	2574		
Qy	594 ccgataaatgtatgaagcatgtgcaagaaagaaacacggaatctctctccactgtttgcaaa	653		
Db	2573 CCATTAAGATTAATTAAGCATGGCAGGAAGAAAGAACAGAAATTTCTCTCAGAGTTGGCAAA	2514		
Qy	654 atctgcagcgtatgaatacttctctcccaactctctctgtgtgataaa--gtacggagca	711		
Db	2513 ATGTCGAGATGAGAATCACTCTTCTCCACCTCTCTGTGGTGGAGCAAGATTAACGGGCC	2454		
Qy	712 ctctcaagggttaatgatctgtctctctctc--caagcggaataatgataaggt--ctccca	768		
Db	2453 ATCTCAGGGTTAATGATTCGGTCTCTCTCTCTCCACAGCGGACTTAATTAATGGCTTCTCAACA	2394		
Qy	769 tgggctgtgcagcgtcat--tccctctcttgaaactgtatgtgccaggtgtgtgtgtataagt	827		
Db	2393 TGGCGCGGTGACGCTCATCTCTTTTTCGCCCTCTGCGCCGACAGGGTGTATTGGATAG	2334		
Qy	828 tatatcatgtgccga-----attccctgtgtatccatgagagctgtgtgaaagcttgtgc	882		
Db	2333 GATGTATTATCATCGCAGGAATTTTAACCGTGAAGTATCATGAGACTTGTCTGA--CGTCTGCC	2276		
Qy	883 ccaatcgtt 892			
Db	2275 CCACATCGTT 2266			

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:31:41; Search time 383.18 Seconds  
(without alignments)  
586.337 Million cell updates/sec

Title: US-08-956-991-8  
Perfect score: 898  
Sequence: 1 tgcgcgcgcgttcgaagcct.....gcccacacatcgttcgtca 898

Scoring table: IDENTITY\_NUC  
Gapcp 10.0, Gapext 1.0

Searched: 311585 seqs, 125096042 residues  
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	898	100.0	898 1 V31986	Mouse Down syndrom
2	502	55.9	6413 1 V31988	Human Down syndrom
3	502	55.9	6604 1 V31981	Human Down syndrom
4	38.8	4.3	1225 1 T60757	Human aortic prefe
5	38.8	4.3	1238 1 V48217	Human aortic-prefe
6	38.8	4.3	2793 1 V48218	Human striated mus
7	37.8	4.2	1308 1 Q25532	Sequence of genom
8	36.2	4.0	1308 1 V48215	Rat aortic prefe
9	36.2	4.0	1308 1 V50485	Streptomyces clavu
10	33.8	3.8	1032 1 V50487	Coding sequence of
11	33.8	3.8	6254 1 V50431	Blood transmissibl
12	33.8	3.8	7193 1 V50431	Synthetic HIV-1 gp
13	33.8	3.8	9436 1 V23291	Protein PEO334 (NP
14	32.6	3.6	2481 1 X52268	Human neural cell
15	32.2	3.6	3997 1 V62739	Protein disciphide
16	32.2	3.5	1555 1 V74310	Human growth facto
17	31.8	3.5	1755 1 Q87670	Human gene - ignatu
18	31.6	3.5	132 1 T27590	Serum somatotropin
19	31.4	3.5	255 1 T24422	H. pylori AlpB gen
20	31.4	3.5	1557 1 V07248	DNA encoding Helic
21	31.4	3.5	1557 1 V07248	Serum somatotropin
22	31.4	3.5	1557 1 V07248	Syngp120nm. Synthe
23	31.4	3.5	1557 1 V07248	HIV-1 gp120 synthe
24	31.4	3.5	1557 1 V07248	Synthetic HIV-1 gp
25	31.4	3.5	1557 1 V07248	Hepatitis C virus
26	31.4	3.5	1557 1 V07248	Syngp100nm. Synthe
27	31.4	3.5	1557 1 V07248	HIV-1 gp160 synthe
28	31.4	3.5	1557 1 V07248	tie truncated rece
29	31.4	3.5	1557 1 V07248	CDNA of Elmeria te
30	31.4	3.5	1557 1 V07248	Elmeria tenella 54
31	31.4	3.5	1557 1 V07248	
32	31.4	3.5	1557 1 V07248	
33	31.4	3.5	1557 1 V07248	
34	31.4	3.5	1557 1 V07248	

C 35	30.8	3.4	2870 1	V71912	S. cerevisiae KIP2
C 36	30.8	3.4	4057 1	T36899	Mouse neuron rest
C 37	30.8	3.4	6567 1	Q00324	Elmeria tenella ge
C 38	30.6	3.4	2001 1	T94200	Thermotoga neapol
C 39	30.6	3.4	4942 1	X00461	Human type VI aden
C 40	30.2	3.4	45546 1	X23520	Human kidney amino
C 41	30	3.3	1563 1	V03867	Mouse hepatoma der
C 42	29.8	3.3	1335 1	X02886	Human mACHR-6 CDNA
C 43	29.8	3.3	2614 1	V48219	Mouse striated mus
C 44	29.8	3.3	2689 1	X02885	Human mACHR-6 CDNA
C 45	29.8	3.3	9578 1	V52260	Streptococcus pneu

ALIGNMENTS

RESULT 1	
ID V31986	standard; cDNA; 898 BP.
AC V31986:	
AT 28-SEP-1998 (first entry)	
DE Mouse Down syndrome-cell adhesion molecule DS-CAM mid cDNA.	
KW DS-CAM; Down syndrome-cell adhesion molecule; neural cell;	
KW signal transduction; trisomy 21; mental retardation;	
KW holoprosencephaly; corpus callosum agenesis;	
KW schizencephaly; diagnosis; assay; mouse; ds; ss.	
OS Mus sp.	
PN M0981795-A1.	
PD 30-APR-1998.	
PF 23-OCT-1997; 019547.	
PR 25-OCT-1996; US-029322.	
PA (CEDA-) CEDARS SINAI MEDICAL CENT.	
PI korenberg JR.	
DI MPI: 98-27191/24.	
PT New isolated Down's Syndrome-cell adhesion molecule - used to	
PT develop products for detection, diagnosis and therapy of	
PT developmental and neurological abnormalities	
PS Claim 2: Page 81; 109p; English.	
CC This cDNA sequence comprises the middle region of a cDNA clone	
CC for murine Down syndrome-cell adhesion molecule (DS-CAM), a	
CC member of a novel subclass of the Ig superfamily with homology to	
CC neural cell adhesion molecules. The 5' region (see V31985)	
CC and 3' region (see V31987) of the clone are also provided. The	
CC murine DS-CAM clone was isolated from a C57 Black/6 mouse brain	
CC cDNA library using human DS-CAM cDNA (see V31981 and V31988) as	
CC probe. The invention also provides human DS-CAM proteins (see	
CC M42086-87), as well as expression vectors and host cells,	
CC transgenic animals, antisense oligonucleotides, and primers useful	
CC for amplification of DS-CAM nucleic acids. DS-CAM polypeptides are	
CC associated with developmental and neurological processes. They can	
CC be used in e.g. neural prosthetic devices used in entubulation	
CC methods of repairing (regenerating) damaged or severed peripheral	
CC nerves. The products can also be used in detection, diagnosis and	
CC therapy of developmental and neurological abnormalities such as	
CC Down syndrome, mental retardation, holoprosencephaly, agenesis of	
CC the corpus callosum, or schizencephaly. Antisense oligonucleotides	
CC are used for inhibiting translation of mRNA.	
Sequence 898 BP; 174 A; 222 C; 245 G; 257 T;	
Query Match	100.0%; Score 898; DB 1; Length 898;
Best Local Similarity	100.0%; Pred. No. 1e-271;
Matches 898; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB 1	tgcgcgcgcgttcgaagccttactacagcgcacacgcgtgaattacagttgtccag 60
DB 1	ttccggccgggttcgaagccttactacagcgcacacgcgtgaattacagttgtccag 60
QY 61	gtgtacacattcgtcgtccggtgtgtcaactatgatgtgtgaactgtgaactcc 120
DB 61	ggtgtacacattcgtcgtccggtgtgtcaactatgatgtgtgaactgtgaactcc 120
QY 121	gtgtcgtacacattcgtcgtccggtgtgtcaactatgatgtgtgaactgtgaactcc 180



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:44 : Search time 216.06 Seconds  
(without alignments)  
497.600 Million cell updates/sec

Title: US-08-956-991-8

Perfect score: 898  
Sequence: 1 tggcggcgggtgtgcaagcct.....gcgccacatcgttcgtcga 898

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/prodata1/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata1/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata1/1na/5C\_COMB.seq:\*  
4: /cgn2\_6/prodata1/1na/5D\_COMB.seq:\*  
5: /cgn2\_6/prodata1/1na/6\_COMB.seq:\*  
6: /cgn2\_6/prodata1/1na/PCTUS9\_COMB.seq:\*  
7: /cgn2\_6/prodata1/1na/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
C 1	38.8	4.3	1225	2	US-08-494-577-11 Sequence 11, Appl
C 2	38.8	4.3	1238	3	US-08-795-868-11 Sequence 11, Appl
C 3	38.8	4.3	2793	3	US-08-795-868-13 Sequence 13, Appl
C 4	36.4	4.1	7218	1	US-08-232-463-14 Sequence 14, Appl
C 5	36.2	4.0	342	2	US-08-494-577-2 Sequence 2, Appl
C 6	36.2	4.0	1308	3	US-08-795-868-2 Sequence 2, Appl
C 7	36.2	4.0	1308	2	US-08-494-577-1 Sequence 1, Appl
C 8	35.2	4.0	1308	3	US-08-795-868-1 Sequence 1, Appl
C 9	31.8	3.5	1755	1	US-08-068-395A-2 Sequence 2, Appl
C 10	31.8	3.5	1755	2	US-08-464-365-2 Sequence 2, Appl
C 11	31	3.5	1632	2	US-08-324-243-34 Sequence 34, Appl
C 12	31	3.5	1632	2	US-08-324-243-34 Sequence 34, Appl
C 13	31	3.5	1632	6	PCT-US95-11511-34 Sequence 34, Appl
C 14	31	3.5	2481	2	US-08-324-243-35 Sequence 35, Appl
C 15	31	3.5	2481	2	US-08-532-390-35 Sequence 35, Appl
C 16	31	3.5	2481	6	PCT-US95-11511-35 Sequence 35, Appl
C 17	31	3.5	3845	4	US-08-220-240A-4 Sequence 4, Appl
C 18	30.8	3.4	2870	2	US-08-468-036-28 Sequence 28, Appl
C 19	30.8	3.4	2870	3	US-08-376-843-28 Sequence 28, Appl
C 20	29.8	3.3	1335	3	US-08-985-090-3 Sequence 3, Appl
C 21	29.8	3.3	2614	3	US-08-795-868-15 Sequence 15, Appl
C 22	29.8	3.3	2689	3	US-08-985-090-1 Sequence 1, Appl
C 23	29.4	3.3	1375	4	US-08-468-812-1 Sequence 1, Appl
C 24	29.4	3.3	6060	1	US-07-551-531-1 Sequence 1, Appl
C 25	29.4	3.3	8252	1	US-08-046-585-15 Sequence 15, Appl
C 26	29.4	3.3	8252	1	US-08-393-703-15 Sequence 15, Appl
C 27	29.4	3.3	8252	6	PCT-US93-11721-15 Sequence 15, Appl

28	29.4	3.3	8937	3	US-08-449-933-1	Sequence 1, Appl
29	29.4	3.3	10706	1	US-08-411-389-1	Sequence 1, Appl
30	29.2	3.3	539	1	US-08-229-515A-11	Sequence 11, Appl
31	29.2	3.3	539	1	US-08-645-865-11	Sequence 11, Appl
32	29.2	3.3	539	6	PCT-US93-06251-8	Sequence 8, Appl
33	29	3.2	1140	1	US-08-289-653-2	Sequence 1, Appl
34	29	3.2	2845	1	US-08-289-653-1	Sequence 1, Appl
35	28.8	3.2	3325	7	5240838-4	Patent No. 5240838
36	28.8	3.2	3624	1	US-07-951-715A-6	Sequence 6, Appl
37	28.8	3.2	3624	3	US-08-459-448A-6	Sequence 6, Appl
38	28.4	3.2	2539	4	US-09-070-060-2	Sequence 2, Appl
39	28.4	3.2	5359	4	US-09-070-060-1	Sequence 1, Appl
40	28.2	3.1	1357	1	US-08-091-519-1	Sequence 1, Appl
41	28.2	3.1	1357	2	US-08-442-043A-1	Sequence 1, Appl
42	28.2	3.1	1357	6	PCT-US91-03478-1	Sequence 1, Appl
43	28.2	3.1	1401	2	US-08-726-136-27	Sequence 27, Appl
44	28.2	3.1	1866	5	US-08-909-742-1	Sequence 1, Appl
45	28	3.1	4157	3	US-08-871-266B-1	Sequence 1, Appl

## ALIGNMENTS

```
RESULT 1
US-08-494-577-11/c
: Sequence 11, Application US/08494577
: Patent No. 5786171
: GENERAL INFORMATION:
: APPLICANT: Lee, Mu-En
: TITLE OF INVENTION: AORTIC PREPERENTIALLY EXPRESSED GENE AND
: TITLE OF INVENTION: AORTIC PREPERENTIALLY EXPRESSED GENE AND
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patcutiln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/494,577
: FILING DATE: 22-JUN-1995
: CLASSIFICATION: 436
: ATTORNEY/AGENT INFORMATION:
: NAME: Fraser, Janis K.
: REGISTRATION NUMBER: 34,819
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1225 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-08-494-577-11

Query Match 4.3% Score 38.8; DB 2; Length 1225;
Best Local Similarity 62.2%; Pred. No. 0.0032;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
592 ccccgatgatgatgatgatgatgcaggaagaaacggaaatctcttcacgttgca 651
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:50 ; Search time 4987.54 Seconds  
(without alignments)  
679.803 Million cell updates/sec

Title: US-08-956-991-8  
Perfect score: 898  
Sequence: 1 tgcgcgcgcgttcgaagcct.....gcgcacacatcgttcgtga 898

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residue  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	58	6.5	605	79	FR0021994	AL014865 F.rubripe
2	58	6.5	618	79	FR0022042	AL014913 F.rubripe

C	3	55.6	6.2	443	79	FR0022007	AL014878 F. rubripes
C	4	52.6	5.9	399	79	FR0022008	AL014879 F. rubripes
C	5	50	5.6	564	79	FR0030722	AL027071 Fugu rubr
C	6	45	5.0	1101	79	CNS00K73	AL077673 Drosophi1
C	7	43.4	4.8	430	79	FR0022072	AL014943 F. rubripes
C	8	43	4.8	419	51	AL1750970	AL005408 Cn08a06. x
C	9	41.8	4.7	405	41	AL1005408	AL005408 Cn08a06. x
C	10	37.8	4.2	388	69	AW114312	AW114312 ts70b05. y
C	11	37.2	4.1	373	34	AA478944	AA478944 zvl8a05. s
C	12	37.2	4.1	433	61	AL181671	AL181961 wj59g07. x
C	13	36.8	4.1	1101	79	CNS00D1D	AL070661 Drosophi1
C	14	36	4.0	583	41	AL1061969	AL061969 MD35081. 3
C	15	35.8	4.0	451	28	AA111432	AA111432 mo29c11. r
C	16	35.6	4.0	470	61	AL181391	AL181391 wj44c10. x
C	17	34.8	3.9	418	39	AL129599	AL129599 LF09781. s
C	18	34.8	3.9	512	30	AA426993	AA426993 LF09781. s
C	19	34.8	3.9	577	37	AA695131	AA695131 GM02190. 5
C	20	34.8	3.9	581	44	AL160382	AL1260382 LP04188. 5
C	21	34.8	3.9	595	45	AL187659	AL187659 GH18241. 5
C	22	34.8	3.9	595	44	AL160812	AL1260812 LP04793. 5
C	23	34.8	3.9	604	42	AL108085	AL108085 GH06818. 5
C	24	34.8	3.9	628	44	AL160736	AL160736 LP04701. 5
C	25	34.8	3.9	645	46	AL105496	AL105496 GH25514. 5
C	26	34.8	3.9	645	46	AL1295438	AL1295438 LP09065. 5
C	27	34.4	3.8	282	59	AV113670	AV113670 AV113670
C	28	34.4	3.8	378	26	W345563	W34563 mc58b05. r.1
C	29	34.2	3.8	220	34	AA4755140	AA475140 wq95c08. r
C	30	34.2	3.8	269	25	W09267	W09267 me05h03. r.1
C	31	34.2	3.8	431	38	AA815874	AA813674 VP.4410. r
C	32	34.2	3.8	446	40	AA982923	AA982923 vx59h05. r
C	33	34.2	3.8	476	27	AA028861	AA028861 mh89d04. r
C	34	34.2	3.8	608	36	AA1511986	AA1511986 vo94b12. r
C	35	34	3.8	648	44	AL135213	AL135213 ms30c11. x
C	36	34	3.8	1159	80	CNS015K9	AL106041 Drosophi1
C	37	33.8	3.8	1023	80	CNS016J3	AL106321 Drosophi1
C	38	33.4	3.7	372	47	AL1059488	AL1059488 vx03c04. y
C	39	33.4	3.7	392	81	B87927	B87927 RBC11-3002
C	40	33.4	3.7	440	40	AL1060692	AL106092 ub41g07. r
C	41	33.4	3.7	448	39	AA874513	AA874513 vx03c04. r
C	42	33.4	3.7	507	74	AA202886	AA202886 fJ2e12. y
C	43	33.4	3.7	569	82	AO709567	AO709567 HS-1484_B
C	44	33.4	3.7	761	62	AI924329	AI924329 wn55e01. x
C	45	33.2	3.7	585	30	AA211275	AA211275 zpa3h08. r

## ALIGNMENTS

RESULT	1
FR0021994	
LOCUS	605 bp DNA
DEFINITION	GSS
ACCESSION	10-DEC-1997
VERSION	
KEYWORDS	F.Rubripes GSS sequence, clone 070J16ab2, genomic survey sequence
SOURCE	AL014865 GI:268123
	GSS: genome survey sequence.
	Fugu rubripes.
	Fugu rubripes.

REFERENCE	1 (bases 1 to 605)
AUTHORS	Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y., Williams,G. and Brenner,S.
TITLE	Direct Submission
JOURNAL	Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: blonhelp@mgp.mrc.ac.uk
COMMENT	Vector: pBluescript II KS

FEATURES

DESCR:	One pass dye-terminator sequencing of cosmid cloned genomic sequence.
location/Qualifiers	

```

source
1..605
/organism="Fugu rubripes"
/db_xref="taxon:11033"
/clone_lib="cosmid 070j16"
/clone="070j16a812"
BASE COUNT      137 a      138 c      154 g      157 t      19 others
ORIGIN

```

Query Match	6.58;	Score 58;	DB 79;	Length 605;
Best Local Similarity	73.0%;	Pred. No. 3.2e-07;		
Matches	73;	Conservative	0;	Mismatches 27; Indels 0; Gaps 0

```

QY      559  ctgcacatgtgattgaa .attccacggtccctcccgtatgaattgatagcatgacag 618
        |||||  |||||  |  |||  |||||  |||||  |||||  |||||
Db      177  ctgcacacatgtacattgatgaccaccaacccctccgatatgattgagcctggcacg 236

```

Oy	619	agaagaacccggaatctctctcactgtgtggcaaatctg	558
Db	237	AGANAAAAACGGAGTCCCTCGCTCAGCTGGTTAAGNTG	276

RESULT	618 bp	DNA	GSS	10-DEC-1997
FR0022042				
LOCUS	FR0022042	618 bp	GSS	10-DEC-1997
DEFINITION	F. rubripes GSS sequence, clone 070M05Ab2, genomic survey sequence.			
ACCESSION	AL014913			
VERSION	AL014913.1	GI:2681281		
KEYWORDS	GSS: genome survey sequence.			
SOURCE	Fugu rubripes.			
ORGANISM	Fugu rubripes.			

**REFERENCE**

**AUTHORS**

1 (bases 1 to 618)

Elgár, G., Clark, M., Smith, S., Meek, S., Warner, S., Umránia, Y., Neopterygii; Teleostei; Euteleostei; Neanchepsopterygii; Tetraodontiformes; Tetraodontoidae; Tetraodontidae; Fugu.

**TITLE** Direct Submission  
**JOURNAL** Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk

V\_type: pnaemaid  
PRIMER: KS  
DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic

FEATURES	location/units
source	1. 618
	/organism="Fugu rubripes"
	of "Fugu rubripes"

BASE COUNT	ORIGIN
160 a	133 c
131 g	165 t
29 others	

[illegible]

QY 260 tgcctcatccgcccgtatgatcctcgltgccttggtcacctcttccaacccgattctt 319  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 277 TGGTCCCTGTACTAATGAGACCTCTTGCTGGAAAACCTCGCGGCCCATTTTATTGTA 336

```
QY      320 ggcgtacatcgcgaagtctgtagtggagaagggttgatcatcaatgaaggtcgccgaatt   379  
          | ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||  
Db     337 NGAACTNCATGCCGAGTGTGAANACCAAGCGGGTCANCCTACTATGTITGCNCTGGTT    396
```

QY	380	cagctgagggaacatccttggtctctcagcagaatcccaagatcgcg	427
Db	397	GTTGGTGGGIGANATNTCTCGTGCTGTACATGTGTCCCAAGAACTTCG	446

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 20:51:16 : Search time 7810.03 Seconds  
(without alignments)  
-844.819 Million cell updates/sec

Title: US-08-956-991-9  
Perfect score: 2173  
Sequence: 1 accaccattcacacaccacg.....aagattgcgaatatatta 2173

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pl4: \*  
11: gb\_pl5: \*  
12: gb\_pl6: \*  
13: gb\_pl7: \*  
14: gb\_pl8: \*  
15: gb\_pl9: \*  
16: gb\_pl10: \*  
17: gb\_pl11: \*  
18: gb\_pl12: \*  
19: gb\_pl13: \*  
20: gb\_pl14: \*  
21: gb\_pl15: \*  
22: gb\_pl16: \*  
23: gb\_pl17: \*  
24: gb\_pl18: \*  
25: gb\_pl19: \*  
26: gb\_pl20: \*  
27: gb\_pl21: \*  
28: gb\_pl22: \*  
29: gb\_pl23: \*  
30: gb\_pl24: \*  
31: gb\_pl25: \*  
32: gb\_pl26: \*  
33: gb\_pl27: \*  
34: gb\_pl28: \*  
35: gb\_pl29: \*  
36: gb\_pl30: \*  
37: gb\_pl31: \*  
38: gb\_pl32: \*  
39: gb\_pl33: \*  
40: gb\_pl34: \*  
41: gb\_pl35: \*  
42: gb\_pl36: \*  
43: gb\_pl37: \*  
44: gb\_pl38: \*

45: gb\_pl39: \*  
46: gb\_pl40: \*  
47: gb\_pl41: \*  
48: gb\_pl42: \*  
49: gb\_pl43: \*  
50: gb\_pl44: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1147.6	52.8	6110	11	AF023449 Homo sapi
2	792.8	36.5	6413	11	AF023450 Homo sapi
3	394.2	18.1	6025	10	AB032958 Homo sapi
4	319.2	14.7	721	11	HUMY291F03 Homo sapi
5	236.8	10.9	1950	12	RNY15054 Rattus norv
6	230.6	10.6	14561	11	AF064862 Homo sapi
7	230.6	10.6	159424	11	AF064865 Homo sapi
8	154.2	7.1	430	13	G36681 SHC-53823
9	109.6	5.0	36156	11	AC004789 Homo sapi
10	109.6	5.0	40619	11	AC005222 Homo sapi
11	109.6	5.0	172042	43	AC012676 Homo sapi
12	81.6	3.8	99699	43	AC013845 Homo sapi
13	51.8	2.4	2658	8	CR078547 Homo sapi
14	47.6	2.2	63751	45	AC017489 Homo sapi
15	47.6	2.2	80446	35	AC004288 Homo sapi
16	45.2	2.1	196667	42	AC012068 Homo sapi
17	44.6	2.1	1853	10	HUM56KDA PR
18	42.4	2.0	7218	5	I66494 Sequence 14
19	42.2	1.9	2250	4	AF069737 Homo sapi
20	41.4	1.9	2094C	43	AC014917 Homo sapi
21	41.2	1.9	2154	35	AF010480 Trypanoso
22	41.2	1.9	32799	8	SPAC227 Arabidops
23	41.2	1.9	91916	8	AC005964 Arabidops
24	41.2	1.9	110680	8	AC006259 Arabidops
25	41.2	1.9	38666	7	SPAC24C9 Arabidops
26	40.8	1.9	35805	8	YSC19576 Homo sapi
27	40.6	1.9	1886	12	AF143957 Mus muscu
28	40.2	1.8	8665	11	HS086136 Human telom
29	39.8	1.8	92586	33	AC009578 Human telom
30	39.2	1.8	1428	5	E14330 cDNA encodi
31	39.2	1.8	1602	10	AB010098 Homo sapi
32	39.2	1.8	3269	9	AB023142 Homo sapi
33	39.2	1.8	46626	4	AF108420 Fugu rbr
34	38.6	1.8	1250	35	AF098070 Drosophill
35	38.6	1.8	1756	35	AF152419 Drosophill
36	38.6	1.8	2884	35	AF117606 Drosophill
37	38.6	1.8	4586	10	HSW801203 Homo sapi
38	38.6	1.8	63433	34	AC004320 Homo sapi
39	38.6	1.8	111336	45	AC018305 Drosophill
40	38.6	1.8	1122	1	MPURSD5T Micromonos
41	38.2	1.8	291288	42	AC008878 Homo sapi
42	38	1.7	2552	34	AF124359 Drosophill
43	37.8	1.7	38406	43	AC013837 Homo sapi
44	37.8	1.7	125013	32	DMBR11L10 Drosophill
45	37.8	1.7	125013	32	DMBR11L10 Drosophill

#### ALIGNMENTS

RESULT 1  
AF023449 6110 bp mRNA  
LOCUS Homo sapiens CHD2-42 Down syndrome cell adhesion molecule (DSCAM)  
DEFINITION MRNA, partial cds.  
ACCESSION AF023449  
VERSION AF023449.1 GI:3169765

[illegible]

ORIGIN

Query Match	52.8%;	Score 1147.6;	DB 11;	Length 6110;
Best Local Similarity	83.7%;	Pred. No. 0;		
Matches 1545;	Conservative	0;	Mismatches 259;	Indels 42;
				Gaps 20;

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 25, 2000, 19:32:03 ; Search time: 383.18 Seconds  
(without alignments)  
1418.830 Million cell updates/sec

Title: US-08-956-991-9

Perfect score: 2173  
Sequence: 1 accaccattcacacacccag.....aagaatgcccaatatatta 2173

Scoring table: IDENTITY-MUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2165.6	99.7	2173	1 V31987	Mouse Down syndrom
2	1147.6	52.8	6604	1 V31981	Human Down syndrom
3	792.8	36.5	6413	1 V31988	Human Down syndrom
4	304.2	14.0	388	1 V31982	Human Down syndrom
5	40.2	1.8	7881	1 V27865	Human telomerase R
6	39.2	1.8	1428	1 V05871	Human neuronal str
7	38.6	1.8	8839	1 V25987	Human telomerase e
8	37.6	1.7	7042	1 V84785	Apoptosis inducer
9	37.6	1.7	7075	1 V84798	Apoptosis inducer
10	36.2	1.7	270	1 V87625	EST clone DY106. N
11	33.4	1.6	1171	1 O13429	Calcium channel ga
12	33.4	1.5	8214	1 V25987	Rat telomerase enc
13	32.8	1.5	585	1 V19497	Cytochrome P45011d
14	32.8	1.5	1494	1 O87730	Human auxillary cy
15	32.8	1.5	1494	1 O87731	Human auxillary cy
16	32.8	1.5	1494	1 O87732	Human auxillary cy
17	32.8	1.5	1494	1 T17388	Human derived cyto
18	32.8	1.5	1494	1 T17388	Human derived cyto
19	32.8	1.5	1494	1 T28395	Human cytochrome P
20	32.8	1.5	1494	1 T28396	Human cytochrome P
21	32.8	1.5	1494	1 T28397	Human cytochrome P
22	32.8	1.5	1494	1 T28398	Human cytochrome P
23	32.8	1.5	1545	1 V19496	Cytochrome P45011d
24	32.8	1.5	2152	1 O70732	TATA-binding prote
25	32.8	1.5	2152	1 T42218	Human TATA-binding
26	32.8	1.5	2152	1 T79594	TATA-binding prote
27	32.6	1.5	1396	1 O98488	Human 50 kDa dystr
28	32.6	1.5	1396	1 T97324	Human (50 kDa) cys
29	32.6	1.5	1436	1 V15734	Alpha-sarcoglycane
30	32.6	1.5	1436	1 V17873	Homo sapiens alpha
31	32.4	1.5	3670	1 O22501	Human MIR gene. Ma
32	32.2	1.5	2085	1 O81504	ORDE 45 kDa subuni
33	32.2	1.5	5496	1 X07327	Aspergillus oryzae
34	32	1.5	1506	1 Q47667	Sequence of the ra

35	32	1.5	1918	1 V54122	Mouse membrane pro
36	32	1.5	2060	1 T71579	Exon 5 of LETO rat
37	32	1.5	2187	1 V61810	Glucocorticoid ind
38	32	1.5	2217	1 X28434	EGF-like homologue
39	32	1.5	2217	1 X52215	Protein PRO230 CDN
40	31.8	1.5	543	1 X52215	Antigen tc-7a gene
41	31.6	1.5	1713	1 X20661	Poly nucleotide seq
42	31.6	1.5	2514	1 T30308	Mouse Sox-9 CDNA.
43	31.6	1.5	3256	1 V29063	BRCA1 modulator pr
44	31.6	1.5	3934	1 T30309	Human SOX-9 CDNA.
45	31.6	1.5	15894	1 V18264	Measles virus Edmo

## ALIGNMENTS

RESULT 1	
V31987	V31987 standard; CDNA: 2173 BP.
AC	V31987.
DT	28-SEP-1998 (first entry)
DE	Mouse Down syndrome-cell adhesion molecule DS-CAM 3' CDNA.
KW	DS-CAM; Down syndrome-cell adhesion molecule; neural cell;
KW	signal transduction; trisomy 21; mental retardation;
KW	holoprosencephaly; corpus callosum agenesis;
KW	schizencephaly; diagnosis; assay; mouse; ds; ss.
OS	Mus sp.
PN	W09817795-A1.
PD	30-APR-1998.
PF	23-OCT-1997; U19547.
PR	25-OCT-1996; US-029322.
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.
PI	Korenberg JR.
DI	WPI: 98-271791/24.
PT	New isolated Down's Syndrome-cell adhesion molecule - used to
PT	develop products for detection, diagnosis and therapy of
PT	developmental and neurological abnormalities
PS	Claim 2; Page 81-83; 109pp; English.
CC	This CDNA sequence comprises the 3' region of a CDNA clone
CC	for murine Down syndrome-cell adhesion molecule (DS-CAM), a
CC	member of a novel subclass of the Ig superfamily with homology to
CC	neural cell adhesion molecules. The middle region (see V31986)
CC	and 5' region (see V31985) of the clone are also provided. The
CC	murine DS-CAM clone was isolated from a C57 Black/6 mouse brain
CC	CDNA library using human DS-CAM CDNA (see V31981 and V31988) as
CC	probe. The invention also provides human DS-CAM proteins (see
CC	W42086-87), as well as expression vectors and host cells.
CC	transgenic animals, antisense oligonucleotides, and primers useful
CC	for amplification of DS-CAM nucleic acids. DS-CAM polypeptides are
CC	associated with developmental and neurological processes. They can
CC	be used in e.g. neural prosthetic devices used in entubulation
CC	methods of repairing (regenerating) damaged or severed peripheral
CC	nerves. The products can also be used in detection, diagnosis and
CC	therapy of developmental and neurological abnormalities such as
CC	Down syndrome, mental retardation, holoprosencephaly, agenesis of
CC	the corpus callosum, or schizencephaly. Antisense oligonucleotides
CC	are used for inhibiting translation of mRNA.
SQ	Sequence 2173 BP; 585 A; 598 C; 554 G; 436 T;
Query Match	99.7%; Score 2166.6; DB 1; Length 2173;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 2169; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
QY	1 accaccattcacacccagacatgagcggttcgcgcgaaccttcagttcgtctcc 60
DB	1 ACCACCTTTACACACCCAGACATGGCGGTTGCGGCAACCTTACGTTCTCGCCTTC 60
QY	61 tgaaggtaaaggctgctgctggtttatagacggcgacatgccatctctgcatacgt 120
DB	61 TGTAGGCTAAAGGCTGCTGCGGTTATAGACCGGCAATGCCATCTCTGCAATACGCT 120
QY	121 ggcacagtggttcacatcgtgattccagcgccaagctaaatctgtctgtgacctg 180

Db	121	GGCCAGGCGGCTTTCACATTCGGATTCCAGGCCCAACCTAAATATCTGTTCTGATGCGCCCTG	180
Oy	181	caagttcaagccgttcacgtctccagctctaaagttcccaagatcgaaagcgtttagatcatagga	240
Db	181	CAGTTTCAGCCGTTTCAGCTCCAGTCTCAAGTTCACATGCGGAAGCGGTTAGATCATAGGA	240
Oy	241	acttgaagccagcatcatalcgcgcagccagggggttgaaagcgcagagaaagttagatattctttctgt	300
Db	241	ACTGGAAACCCAGTACATTCGCGCAGCGAGGGGTGGAAAGCGCAAGTAGATCTTTTCTGT	300
Oy	301	gtgagccgttgagcccaagctcagggttggttgaaagaccttcgcagccagcgagcgttcat	360
Db	301	GTGGCTGTGAGCACAGTCTCAGGTGTGTGAAGACATTTCTGAGCCAGGAGGGTTTCAT	360
Oy	361	accgcttgaaaaccctaaagttgtggagcttgataaagttcacccctacgtgccaaaatg	420
Db	361	ACCGCTTGAAACCTAAAGTGTGGAGCTTGATTAATGTTACCTTCAACCTTCACTCCCAAAATG	420
Oy	421	gaagtagtcccgggcgccataagtgaaatcatagaaagccaatgcccttgaggaaagaaacccc	480
Db	421	GAGTAGGTCCCGGGCGCATAAAGTGAATCATAGAAACCAAAACCTGGGGAAGAAACCC	480
Oy	481	agttctccaagagcagagagcttttcgcgcagcatcaataacacccagtlvaggctgaatc	540
Db	481	AGTTCTCCAAGAGCAGAGAGCTTTTCGCAGCATCAATACCCCGAGTAGAGGCTGAATC	540
Oy	541	tgaattgcttgaaaagcagggcggtctgccaatctactctactcttcaatcagacacct	600
Db	541	TGAAATGCTGGAATACAGCGCGGCTGTCCAAATACCTCATTCCTTGAAATCAACACCT	600
Oy	601	ttggagcaacaggtcgcgcagcagcttcaagcgagacctcccttcccaagtlcctaacaaltctg	660
Db	601	TTGGCAACACGGGTCTGAGCACAGCTCAAGCGACCTTCTTCCAAAGTCTCAACATTTCTG	660
Oy	661	taagaccgttgaaagcaagcagtggtatgaactgcgagataagaggtgagcaagagcgccggcg	720
Db	661	TATGACCTTCAGAAACCCAGTGTGATGAATGCAATAGATAGATGATGTCAAACAGCGCCGC	720
Oy	721	tgtgaggaataagaaagccaacttcgcacagctgaactagatagtcgacatgaatcccttcca	780
Db	721	TGTGGCGATAGCAACCCCAATCTTGGCACGCTCACTAGATGTGGAGTGCMAATCCCTCCA	780
Oy	781	ctcatctaagtcadttgtccacaagcgaaagaggtcgtgacacccaacgaagggcttaaga	840
Db	781	CTCATTAAGTCACATTTGTCCACAAGCAAGAAAGCGCTGCACAAACGAAGGGCTCAAGA	840
Oy	841	tcctcgtgaccatctcctcgtcatccctggttcggggtttctacgtcccttgagcttcgtcg	900
Db	841	TCCCTCGTAGACATCTCTCGCATCTCTGTCGGGTTCTACTGCTCTTTGTCTCTCTCTCG	900
Oy	901	ttgtgcgagagagacgcgcagagcagagagctctgaagagcgctgaagagatgycaaagatttag	960
Db	901	TTGTGCGGAGAGACGGCGGAGCAGAGAGGCTGAAGAGCGCTGAGAGATGCAAAAGATTTAG	960
Oy	961	ctgaagatgctcatgagcaaaaacacacacgcgagcttcaggtlactcttaagaaacacagcagaga	1020
Db	961	CTGAATGCTCATGAGCAAAAACACACAGGACTTCAGATACCTTTAAGCAAAACAGCAGCAGA	1020
Oy	1021	ctttgaaagatgacatttgatatataaccaagaggtctagctcttgattgaagagagagacacaa	1080
Db	1021	CTTTGAGAAATGCACATTGATATATACCCAGGCGCTAGTTTGTGATGAGAGAGACACAA	1080
Oy	1081	tgagagaccataagatgagccgctcccaagttcctgttgacggatgctgacttcggggagagcag	1140
Db	1081	TGAGAGACCATGATGAGACCGCTCCACAGTCTCTGTGAGGAGATGCTGACTTCGGGAGGCGAG	1140
Oy	1141	ccaaacgagaaagttcctgacagttgagctcaacagagtgacattaccaatctggtgtcctaggcca	1200
Db	1141	CCAAACGAAATGCTACTGACAGTGACTACACAGGTTGTTATACCATTCGGTGTCTAGGCGCA	1200
Oy	1201	ccggagccctctgtagatltccctcgatgtctgcgcaagaaacgaatcccacccagagaga	1260

Db	1201	CCGGGCCCCCTCGTGGATGTCTCCGATGCTCGGCCAGGATGAATCCACCACGACGAGGA	1260
Qy	1261	atgcaaaagctcggagcccaacagcgaagaaccgtaacgcacagctggagcgtcaaacagac	1320
Db	1261	ATGCAAAAGCTTGAGCCCAACGAGAAACCGGTAACGGCAGCACTGGACGCTCAACAGAC	1320
Qy	1321	cccatcctaccatctctcgacacacacctcacacagaatlgagactgttacccaaggtac	1380
Db	1321	CCCATCTCAACCTACTCTCTGCACACACCCCTCACACAGATGAGACTGTCTACACAGGCTAC	1380
Qy	1381	aggatccgttgaaagaaagagcgacatagacgctacagcccatcacaagaacagacagcagc	1440
Db	1381	AGGATCCGTAAGAGAGAGAGCGCACACTACAGGCTAGCCCATTCACAAAGACACAGACGAC	1440
Qy	1441	aagaagcagcactgtctccacagaaagtgtctcttactactagaagaactcgcagacctta	1500
Db	1441	AAGAAGCAGCACTGTCTCTCCACGAAAGTGCTCTTCTTACTACGAAAGCTCCAGGGCTTA	1500
Qy	1501	tgaacacgcgaagatcggaaagagcacttgaagcatgcgaagctcccaagctcacagagtctt	1560
Db	1501	TGAACACGCGCAAGATGAGAGAGACACTGAGGCAATGCCAAGTTTCACATCACAGAGTGTT	1560
Qy	1561	catatccgaatacgtccctccgagcagcttgacgycgaagacaatatgagttacaggaacgtctg	1620
Db	1561	CATATCCGATAGGTCTCTCCGACAGTGAACGCGACAGCAAAATGATACACGACAGTGTG	1620
Qy	1621	actccagttacctcttaagaatacggagatcgcagatcttaagctatctgcactctcccccacactcag	1680
Db	1621	ACTCCAGTACCCCTTTAGAATGGGATGTGCGAGATTATGACTATCTCCCCCAACTCTAGG	1680
Qy	1681	atggaagacagctgtgaacaatgacgagcttcccaagaagcccatcgcgcgaagcgactacac	1740
Db	1681	ATGGAAGACAGTGTGAACATGAGATGGCGGTTCCAAAGGCCCATCGGCCAGGGAGACTCTACAC	1740
Qy	1741	ctgcctccatacctcagatggtatctctgttaaacccggcgccacgaagcaccagcagagctg	1800
Db	1741	CTGCTCATACCTACGATGATGATTCTTTGAACCCGGGCGCACAGGCACACAGACTAC	1800
Qy	1801	agcttgaagacaagcgtgctctggaaccccaagaagctcgaacctgaaacgcgcccaagctcg	1860
Db	1801	AGTTTGAAGACAAAGCGTCTTGGAACCCCAAGAAAGTCGGACCTCGAAGCCGCCACGGGTG	1860
Qy	1861	ttagagccacacctatggaagcctcctctcctcaactctctccacgcgagaagaagagcagct	1920
Db	1861	TTGAGCCACACCCCTATGAGAGCGCTCCCTCCCTCCACTTCTTCCACGCGAGAAGGACACACT	1920
Qy	1921	cgttagcaacaagagctgtgtgacacaccttaccacgaagagaggtgcagagctcgagacagcg	1980
Db	1921	CGTGGCAACAAGGGGCTGTGGCCACTTACTACCGAAGAGGGTGCAGAGCTGGACAGCG	1980
Qy	1981	agctlaaaatgagcagctccccaagaagctacgtctgagctcccgaggcaattgaaagagaca	2040
Db	1981	AGCTAAATGAGCGAGCTCCCAAGAGTCACTGAGATCCCGGGCAATTGAAAGAACAA	2040
Qy	2041	tcctcagcaaatcttaacactctgtgataacaacatgacatgacatgacatgacacgagttgaat	2100
Db	2041	TCCTCACCAATCTTACACTTGTGTTAACACATGCACTGATGACACCGGTTGAAAT	2100
Qy	2101	acaattacgcagccaatcaatgaactctcttattgaaattccgaattattaaattagaact	2160
Db	2101	ACAATTACGAGCCCAATCAACTACTTTTATGAAATTCGATATTATTAATTAAAGAAAT	2160
Qy	2161	gccaaatatatta 2173	
Db	2161	GCCAAATATATTA 2173	
RESULT 2			
ID	V31981	standard: cDNA: 6604 bp.	
AC	V31981:		
DT	28-SEP-1998	(first entry)	
DE	Human Down. vndrome-cell	adhesion molecule DS-CAM1 cDNA.	

RESULT	2
ID	V31981
	V31981 standard; cDNA; 6604 BP.
AC	V31981;
DT	28-SEP-1998 (first entry)
DE	Human Down syndrome-cell adhesion molecule DS-CAM1 cDNA

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 19:24:56 ; Search time 216.06 seconds  
(without alignments)  
1204.103 Million cell updates/sec

Title: us-08-956-991-9  
Perfect score: 2173  
Sequence: 1 accaccatccaccaccag.....aagaatgcccaatatatta 2173

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 211294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428568

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/5C.COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/5D.COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/6.COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/PCTUS9\_C/4B.seq: \*  
7: /cgn2\_6/prodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.4	2.0	7218	1	US-08-232-463-14 Sequence 14, Appl
C 2	40.2	1.8	7881	4	US-08-751-189-1 Sequence 1, Appl
C 3	40.2	1.8	7881	4	US-09-060-836-1 Sequence 1, Appl
C 4	33.8	1.6	1171	2	US-08-336-257A-1 Patent No. 5386025
C 5	33.8	1.6	1171	7	5386025-1 Sequence 17, Appl
C 6	32.8	1.5	2152	1	US-08-188-582-17 Sequence 17, Appl
C 7	32.8	1.5	2152	1	US-08-646-715-17 Sequence 11, Appl
C 8	32.6	1.5	1396	1	US-08-483-278-11 Sequence 8, Appl
C 9	32.6	1.5	1396	1	US-08-283-917-8 Sequence 8, Appl
C 10	32.2	1.5	2085	3	US-08-961-716-8 Sequence 13, Appl
C 11	32.2	1.5	2085	3	US-07-937-609-13 Sequence 13, Appl
C 12	32.2	1.5	533	7	5482709-5 Patent No. 5273901
C 13	31.8	1.5	543	7	5482709-6 Patent No. 5273901
C 14	31.8	1.5	3256	4	US-08-968-751-3 Sequence 3, Appl
C 15	31.6	1.5	3256	4	US-08-627-873-6 Sequence 6, Appl
C 16	31.4	1.4	3640	2	US-08-472-217-1 Sequence 1, Appl
C 17	31.4	1.4	26700	2	US-08-488-199-5 Sequence 5, Appl
C 18	31.4	1.4	26700	4	US-08-145-658D-23 Sequence 23, Appl
C 19	31.2	1.4	1569	4	US-08-368-803-16 Sequence 16, Appl
C 20	31.2	1.4	2521	2	US-08-578-096A-18 Sequence 18, Appl
C 21	31.1	1.4	2521	4	US-08-599-252-81 Sequence 81, Appl
C 22	30.8	1.4	1050	2	US-08-436-074-54 Sequence 54, Appl
C 23	30.8	1.4	1050	2	US-08-596-06352-81 Sequence 81, Appl
C 24	30.8	1.4	1050	6	PCT-US96-06583-81 Sequence 8, Appl
C 25	30.6	1.4	2149	2	US-08-784-651-3 Sequence 3, Appl
C 26	30.6	1.4	2149	2	US-08-784-651-3 Sequence 3, Appl
C 27	30.4	1.4	1365	2	US-08-420-235B-32 Sequence 32, Appl

C 28	30.4	1.4	1365	6	PCT-US95-10194-32 Sequence 32, Appl
C 29	30.4	1.4	1501	4	US-08-145-658D-24 Sequence 24, Appl
C 30	30.4	1.4	1566	4	US-08-145-658D-13 Sequence 13, Appl
C 31	30.4	1.4	1701	1	US-07-863-169A-2 Sequence 2, Appl
C 32	30.4	1.4	1701	4	US-08-429-96A-2 Sequence 2, Appl
C 33	30.4	1.4	1701	6	PCT-US93-08062-2 Sequence 2, Appl
C 34	30.4	1.4	20710	6	US-08-420-235B-1 Sequence 1, Appl
C 35	30.4	1.4	20710	6	PCT-US95-10194-1 Sequence 1, Appl
C 36	30.4	1.4	35100	3	US-08-770-379-18 Sequence 20, Appl
C 37	30.2	1.4	1568	4	US-08-145-658D-20 Sequence 18, Appl
C 38	30	1.4	1563	4	US-08-743-637B-183 Sequence 183, Appl
C 39	30	1.4	2307	3	US-08-967-101-28 Sequence 28, Appl
C 40	30	1.4	2307	4	US-08-592-541-28 Sequence 60, Appl
C 41	30	1.4	2416	2	US-08-592-126-60 Sequence 61, Appl
C 42	30	1.4	2416	2	US-08-592-126-61 Sequence 63, Appl
C 43	30	1.4	2416	2	US-08-592-126-63 Sequence 1, Appl
C 44	30	1.4	3366	3	US-08-469-802B-1 Sequence 1, Appl
C 45	30	1.4	3366	3	US-08-267-803B-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFFINGER, F.  
APPLICANT: FALKNER, F.G.  
TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-F15  
US-08-232-463-14

```

Query Match      1.88;  Score 40.2;  DB 4;  Length 7881;
Best Local Similarity 52.7%;  Pred. No. 0.033;
Matches      87;  Conservative      0;  Mismatches      78;  Indels      0;  Gaps      0.

```

QY 104 ccatctgtgcatacggatgagccagtggcgttccattcgatttcgaagccaagttaaactc 16  
Dd 5342 CTTCCCAACGACAACGGTGTAGTCAGCAGCGGCAGCTCTGGGCTCAGGCCAGCACGAATGTATT 528  
OY 164 tttctctgatggcccttgcaagtttagacggttcagcttcgaagttcccagaatgcga 22  
Db 5282 TGGTAAGTGATGGAGCCCTTAGTGTGACGACCACCCGAAACCATGACTGTCCAGGATCCCAGAAGCTCC 52  
OY 224 acggttatgatcatagaaacttgaaaccagatcacatcagtcaggcagccag 268  
Db 5222 AGGAGCCCCGTTCGAGAGGCAGTAAGAAAGAGTGTATCATCTGGAGAGG 5178

RESULT 3  
US-09-060-836-1/c  
Sequence 1, Application US/09060836  
Patent No. 5981707  
GENERAL INFORMATION:  
APPLICANT: Harrington, Lea A.  
TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen, Inc.  
STREET: 1840 De Havilland Drive  
City: Thousand Oaks  
STATE: California

```

1      RESULT 3
2      US-09-060-836-1/C
3      ; Sequence 1, Application US/09060836
4      ; Patent No. 5981707
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Harrington, Lea A.
7      ; APPLICANT: Robinson, Murray O.
8      ; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
9      ; TITLE OF INVENTION: 1
10     ; NUMBER OF SEQUENCES: 12
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Amgen, Inc.
13     ; STREET: 1840 De Havilland Drive
14     ; CITY: Thousand Oaks
15     ; STATE: California
16     ; COUNTRY: USA
17     ; ZIP: 91320-1789
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/09/060.836
25     ; FILING DATE:
26     ; CLASSIFICATION:
27     ; PRIOR APPLICATION DATA:
28     ; APPLICATION NUMBER: US 08/751,189
29     ; FILING DATE: 15-NOV-1996
30     ; ATTORNEY/AGENT INFORMATION:
31     ; NAME: Oleski, Nancy A.
32     ; REGISTRATION NUMBER: 34,688
33     ; REFERENCE/DOCKET NUMBER: A-433
34     ; INFORMATION FOR SEQ ID NO. 1:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 7881 base pairs
37     ; TYPE: nucleic acid
38     ; STRANDEDNESS: single
39     ; TOPOLOGY: linear
40     ; MOLECULE TYPE: cDNA
41     ; US-09-060-836-1

```

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 16:58:56 ; Search time 4987.54 Seconds  
(without alignments)  
1645.003 Million cell updates/sec

Title: US-08-956-991-9  
Perfect score: 2173  
Sequence: 1 accaccattcacacacccag.....aagaattgccaatatatat 2173

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	292.2	13.4	541	64	AW046213
2	239	11.0	310	21	FL3426 HSC2XA021 n



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2000, 21:06:07 ; Search time 7810.03 Seconds

(without alignments)  
-2493.247 Million cell updates/sec

Title: US-08-956-991-10

Perfect score: 6413  
Sequence: 1 tgcacgagcgccgagcagcgcg.....gaaatgccaataatatt 6413

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_ov:\*  
23: em\_ov:\*  
24: em\_pal:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_hlg1:\*  
33: gb\_hlg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pl4:\*  
41: gb\_hlg3:\*  
42: gb\_hlg4:\*  
43: gb\_hlg5:\*  
44: gb\_hlg6:\*

45: gb\_hlg7:\*  
46: em\_hlg1:\*  
47: em\_hlg2:\*  
48: em\_hlg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	6413	100.0	6413	11 AF023450	AF023450 Homo sapi
2	5718	89.2	6110	11 AF023449	AF023449 Homo sapi
3	1845.8	28.8	6025	10 AB032958	AB032958 Homo sapi
4	699	10.9	721	11 HUMV291F03	HUMV291F03 Homo sapi
5	495.4	7.7	109866	11 AF043945	AF043945 Homo sapi
6	469.4	7.3	145861	11 AF064862	AF064862 Homo sapi
7	358.8	5.6	430	13 G36681	G36681 SHGC-53823
8	319.4	5.0	119678	11 AF064866	AF064866 Homo sapi
9	319.4	5.0	120007	11 AF064864	AF064864 Homo sapi
10	307.2	4.8	121019	11 AF042091	AF042091 Homo sapi
11	307.2	4.8	155407	11 AF042090	AF042090 Homo sapi
12	267.4	4.2	159424	11 AF064865	AF064865 Homo sapi
13	154.6	2.4	189607	32 AP000757	AP000757 Homo sapi
14	154.6	2.4	223724	32 AP000834	AP000834 Homo sapi
15	153.8	2.4	193789	32 AP000849	AP000849 Homo sapi
16	153.8	2.4	124395	44 AC013397	AC013397 Homo sapi
17	121.4	1.9	193789	32 AP000849	AP000849 Homo sapi
18	98	1.5	96870	41 AC009739	AC009739 Drosophila
19	96.4	1.5	279137	41 AC008540	AC008540 Homo sapi
20	91.8	1.4	110000	45 AC017014_0	AC017014_0 Homo sapi
21	91.2	1.4	174707	45 AC017082	AC017082 Homo sapi
22	89.2	1.4	107475	45 AC017106	AC017106 Homo sapi
23	89	1.4	7606	34 DROLARW	DROLARW D. melanogaster
24	89	1.4	31646	45 AC017967	AC017967 Drosophila
25	89	1.4	90563	42 AC010115	AC010115 Drosophila
26	88.4	1.4	110000	45 AC017014_1	AC017014_1 Homo sapi
27	87.2	1.4	87042	43 AC011885	AC011885 Homo sapi
28	86.6	1.4	20940	43 AC014917	AC014917 Drosophila
29	86.6	1.4	110000	45 AC017014_1	AC017014_1 Homo sapi
30	86.6	1.4	141025	44 AC015693	AC015693 Homo sapi
31	86.4	1.3	181634	44 AC016253	AC016253 Homo sapi
32	86	1.3	96106	45 AC017015	AC017015 Homo sapi
33	83.4	1.3	162448	43 AC010827	AC010827 Homo sapi
34	83.2	1.3	147343	44 AC011871	AC011871 Homo sapi
35	83.2	1.3	173721	43 AC015866	AC015866 Homo sapi
36	83	1.3	107475	45 AC017106	AC017106 Homo sapi
37	82.8	1.3	840	8 CNSO1BNS	ALL14464 Botrytis
38	81.4	1.3	56416	41 AC011152	AC011152 Homo sapi
39	80.4	1.3	182110	44 AC005302	AC005302 Mus musculus
40	80.4	1.3	194355	33 AL133216	AL133216 Homo sapi
41	80.2	1.3	139207	41 AC011429	AC011429 Homo sapi
42	80.2	1.3	159836	44 AC008334	AC008334 Drosophila
43	79.6	1.2	181634	44 AC016253	AC016253 Homo sapi
44	79.4	1.2	60639	43 AC015463	AC015463 Homo sapi
45	79.4	1.2	67220	45 AC017014_3	Continuation (4 of

## ALIGNMENTS

RESULT 1	AF023450	6413 bp	mrna	PRI	01-JUN-1998
LOCUS	AF023450				
DEFINITION	Homo sapiens CHD2-52	Down syndrome cell	adhesion molecule (DSCAM)		
ACCESSION	AF023450		complete cds.		
VERSION	AF023450.1	GT:3169767			

KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
AUTHORS Eukarya: Primates: Catarrhini: Homiidae: Homo.  
TITLE 1 (bases 1 to 6413)  
Yamakawa, K., Huo, Y.-K., Haendel, M. A., Hubert, R., Chen, X.-N.,  
Lyons, G. E., and Korenberg, J. R.  
DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a  
Down Syndrome Region and is Involved in the Development of the  
Nervous System  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6413)  
AUTHORS Yamakawa, K., Huo, Y.-K., Haendel, M. A., Hubert, R., Chen, X.-N.,  
Lyons, G. E., and Korenberg, J. R.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research  
Institute, 110 George Burns Road, Davis Building, Suite 2005, Los  
Angeles, CA 90048-1869, USA  
FEATURES  
source  
1..6413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21 (trisomy 21)"  
/map="21q22, between HMG14 and MX1"  
/tissue\_type="brain"  
/dev\_stage="14 weeks, fetal"  
/clone="CHD2-52"  
/note="derived from alternately-spliced mRNA"  
1..6413  
/gene="DSCAM"  
/gene="DSCAM"  
453..5168  
/note="member of immunoglobulin superfamily; involved in  
nervous system development"  
/codon\_start=1  
/product="down syndrome cell adhesion molecule"  
/protein\_id="GI:3169768"  
/db\_xref="GI:3169768"  
/translation="MILALSLFOSFANVSEDIHSLYFVNASLOEVPFASTGTLY  
PCPAGIPVTLRWLATGEETIDPGRIHVPNGTQIIPPPSSFTLIHNTYYC  
TAENPSGIRSDQVHIAKALREPTVREDQKTRGNVAFKCIIPSSVAYITVISM  
EDTVSLVSGREFLITGALYIKDVNEDGLYRCITRRTYGTERTOSARLYS  
DANASPSILDFDHRKAMAGORELPCKALHPEPDYRWLKDMPLELSGRQKVT  
GLIENIRPSDGSVYCEVSNRGTAKVIGRLYKOPALATISPRKXSSGVSYLS  
CSVTGDEDELSTWANGELINDGKNVITGINHEMLIDHWKSDGAYCCFYKDL  
SAODIVOVVLEDCGPKIISATSEKVSAPBVSIMCNVKTPLITITWIDDDPLKG  
GSHRISOMITSEGNVSYLNTISSQVRDGGYRGTANNAGVLYOARINVRPASIR  
PKMNIITAIAGRDYIHCRTGYPIYSIKWYKNSLLPNNHQAFFENNGTILKSDYK  
EVEDEGTCTNVLPOLSTQSVHVTVPPIOPFERPRSTICORVIFCVYSDDL  
PLITITWOKDGRPLPGSLGVTINDIDFTSLRLSNLSMHNNGYICIANEAAVAEHOS  
OLIVRPKPVOPRDOGIYKAVILKNSAGYVPTIWKFSKGAQVDFOPDIALN  
CLIRYSPKFNKHYVEDSGYLCKNSDNGADVSKSLTVKIPAMISYPTNL  
GRIOVSKNSLKHVYEDSGYLCKNSDNGADVSKSLTVKIPAMISYPTNL  
ATGOKKEMSTAHGKPLIVWEKEDRIIPENMAYLVSRTKEVEEELISLOLLPY  
REDSGFFSCHAIINSYEDGRITQIVLQEPDPEPELIDVYARTITLMTGFGNSP  
ITGYDECKNKSQSDSAQRTKQVSPQLSATITDIHSSYTSIMYAKNKGISFES  
NELITIADEAAPDPGPOEVHLEPISOSIRVWPKPKHKLONGIKRQITREYSIG  
GNCFINISVDTSQDEVTLDNLKFTQYGLVACNRAGTGPSSQEIITTELEDP  
SYPPENVOAATPESISISMSILKEALNGILQGRVYIYANLMDGLGKIKNITT  
OFSLELDGAEKTYNTSIQVLAFTBAGDVSFOIFTRKEDVGPAPVAKAASAM  
VAVSWLPEPLKNGIIRKTYVFCGHPPTVIFSEFASPSFSEIRIPNLRNOYSWVY  
AVTSAGNSSEIITVEPLAKARILITSGVTYTPMKMDIYLPCKAAGDSPILKVM  
KDSNGTSLVITIDGRSIFSNFSFIIIRTKAEDSGYISCIANNKNSPEIILNOYV  
KDSNGTSLVITIDGRSIFSNFSFIIIRTKAEDSGYISCIANNKNSPEIILNOYV  
RLEKLCGTWKFTILTAONGVGRISSEIIETKTLQEKQFSEKDELFASITVRL  
NLIGMNGCCPITSFTELEYFGTITWTAORTLSKSYIILDOEATWELQMWVCN  
SAGCAEQAQEAARCKEFS"

BASE COUNT 1633 a 1779 c 1709 g 1292 t  
ORIGIN

Query Match 100.0%; Score 6413; DB 11; Length 6413;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 6413: Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Oy	1	tgactgaagccggaagacacggaagaatgagcctgcgcccgcctgtgtcgtgataccg	60						
Db	1	TGACTGAGCCGGAGACAGCGAAAGATGAGCTTGCCTCCGCTGCTGCTGATCCG	60						
Oy	61	aggatgaagggctggcgcaaggagacgctgctgcgtcattcttgagcgccagatgccg	120						
Db	61	AGGATGAGGGCTGGCGACAGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG	120						
Oy	121	gatgaagcagcgcgcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt	180						
Db	121	GATGAGCTACGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	180						
Oy	181	agctgaagcgcgcgcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt	240						
Db	181	AGCTGAGCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	240						
Oy	241	gcaaggaagcagc	300						
Db	241	GCAGAGCGCAGCG	300						
Oy	301	ccgagc	360						
Db	301	CCGGGCGGCTAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	360						
Oy	361	caggctgcagc	420						
Db	361	CAGGTGCCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	420						
Oy	421	tcgctcgc	480						
Db	421	TCCGCGCGGAGCG	480						
Oy	481	agagctgcgaatgtttcagtgtaagacctaactcagcctcattctgtcaatgct	540						
Db	481	AGAGTTGGCGAATGTTTTCAGTGAAAGACTACACTCCACTTCTACTTGTCAATGCT	540						
Oy	541	cctcgaagaggtatgtttcgc	600						
Db	541	CTCTCGAAGAGGTATGTTTGCAGCACACACAGGGGACTCTGCTGCTGCTGCTGCTG	600						
Oy	601	gcatccctcctgtcgtcctcagatgtaactcagcgcgcgcgcgcgcgcgcgcgcgcgcgc	660						
Db	601	GCACTCCTCTGTACTCTCAGATGATGATGATGATGATGATGATGATGATGATGATG	660						
Oy	661	ccggagtcgc	720						
Db	661	CCGGATCCGCGCAGCG	720						
Oy	721	gctcagtaaccttaactcagtaactaactaactaactaactaactaactaactaactaact	780						
Db	721	GCTCTAGTACTTAATCTCATGATATATATATATATATATATATATATATATATAT	780						
Oy	781	aaattgaagtcagagatgtccacatcaagagcgttttcgggagcccatcaactcgtg	840						
Db	781	AAATTAAGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	840						
Oy	841	tggagagcccaagaaacatgagagaggaatgttcggtcctcaagtgcattatccctct	900						
Db	841	TGGAGGACCAAGAAACATGAGAGGAAAGTTCGGCTCTCAAGTGATATATATATAT	900						
Oy	901	cggtgagc	960						
Db	901	CGGTGAGCGGCTACATCAGTGTCTCATGAGAGAAACACTTTCACCTGTCTGAG	960						
Oy	961	gactcaattctcattcattcattcattcattcattcattcattcattcattcattcattc	1020						
Db	961	GATCTAATTTCTATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT	1020						
Oy	1021	atgattgtataactacgcctgcatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1080						
Db	1021	ATGATTGTATAACTACGCTGCATCAGCGCGCATCGATACACCGAGAGAGAGAGAG	1080						



[illegible]

QY	1081	gcaacagcgccagactttttgtatcaagaccacggaactcagcccatcatcttgatg	1140
QY	1081	gcaacagcgccagactttttgtatcaagaccacggaactcagcccatcatcttgatg	1140
QY	1141	ggtcttaccatctcaagcacaatggtcttgacagcgctgttgaaactctcttgcaagcgctcg	1200
QY	1141	ggtcttaccatctcaagcacaatggtcttgacagcgctgttgaaactctcttgcaagcgctcg	1200
QY	1201	ggcaaccttgagccagattcaacgcttggtctgaagaaacatgccccttgaaactttcaaggga	1260
QY	1201	ggcaaccttgagccagattcaacgcttggtctgaagaaacatgccccttgaaactttcaaggga	1260
QY	1261	ggttccaagaagaccggtgaagggagcgtggtcatttggaatactggcccttgagactcaagca	1320
QY	1261	ggttccaagaagaccggtgaagggagcgtggtcatttggaatactggcccttgagactcaagca	1320
QY	1321	gctatgtcttgaaagtgtccaacacagatacgaacactgctcaagggtgatagagcgccctgaag	1380
QY	1321	gctatgtcttgaaagtgtccaacacagatacgaacactgctcaagggtgatagagcgccctgaag	1380
QY	1381	tgaaacagccacttgaaagccaccatcagctcccaagaaagttaaagaagaggttgatgacc	1440
QY	1381	tgaaacagccacttgaaagccaccatcagctcccaagaaagttaaagaagaggttgatgacc	1440
QY	1441	aagttctctgtctctcagcggtgacagagaaactgaagaccagaacctctctgtatcgca	1500
QY	1441	aagttctctgtctctcagcggtgacagagaaactgaagaccagaacctctctgtatcgca	1500
QY	1501	atgggtgaatcctcaacccttgaaaaaaatgtgaggaatacaagagatacaacaaagaaaac	1560
QY	1501	atgggtgaatcctcaacccttgaaaaaaatgtgaggaatacaagagatacaacaaagaaaac	1560
QY	1561	ttataatgatacaatggtccaagaatgaacggggagcgcatatacagtgctcttgcgcaag	1620
QY	1561	ttataatgatacaatggtccaagaatgaacggggagcgcatatacagtgctcttgcgcaag	1620
QY	1621	acaagctgtcgcgtccaagactatgtgcaagtggtccttgaagatggaactcccaaatla	1680
QY	1621	acaagctgtcgcgtccaagactatgtgcaagtggtccttgaagatggaactcccaaatla	1680
QY	1661	ttcttgaccttagtggaaaaggtgtgaagtcagagaaacggtttcccttatgtgcaag	1740
QY	1661	ttcttgaccttagtggaaaaggtgtgaagtcagagaaacggtttcccttatgtgcaag	1740
QY	1741	tgaaaggaaacacctttgcccacatcaactgaaccttgagacatgaaccgatalctcaag	1800
QY	1741	tgaaaggaaacacctttgcccacatcaactgaaccttgagacatgaaccgatalctcaag	1800
QY	1801	gtggcagtcacgcgcatcaaccuaatgattcaagtcwgaagggaggtggttgactacttga	1860
QY	1801	gtggcagtcacgcgcatcaaccuaatgattcaagtcwgaagggaggtggttgactacttga	1860
QY	1861	acatctccaagctcccaaggtcggagacggggaggtctaacgcgtgcacgtcccaacaactcgg	1920
QY	1861	acatctccaagctcccaaggtcggagacggggaggtctaacgcgtgcacgtcccaacaactcgg	1920
QY	1921	cggagatgctctctgtacaaagctcgataaactgaagaggctgcgcgaagcatttgaccaa	1980
QY	1921	cggagatgctctctgtacaaagctcgataaactgaagaggctgcgcgaagcatttgaccaa	1980
QY	1981	tgaaaaacaatacagcagaatgacagagcggcagacacataatactgctcgttgattgct	2040
QY	1981	tgaaaaacaatacagcagaatgacagagcggcagacacataatactgctcgttgattgct	2040
QY	2041	atccgtatctatcatatlaaattgglacagaactctaaccgtcttcccttcaacacgcgc	2100
QY	2041	atccgtatctatcatatlaaattgglacagaactctaaccgtcttcccttcaacacgcgc	2100
QY	2101	aagtgacttttgaaaaaactgaagaaactttaaactttcaatgttgaaaaaagaagatggagcg	2160
QY	2101	aagtgacttttgaaaaaactgaagaaactttaaactttcaatgttgaaaaaagaagatggagcg	2160
QY	2161	aaaggagatacagtgcaacggtgtgtgttcaacacaaacttccacacagcgagagctgc	2220







\*\*\*\*\*  
 WIRE (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Apr 25 13:12:33 2000: MasPar time 43.18 Seconds  
 861.849 Million cell updates/sec  
 Tabular output not generated.

Title: >US-08-956-991-11  
 Description: (1-1571) from US08956991A.pep  
 Perfect Score: 11189  
 Sequence: 1 MWILASLFOSFANVFSEDL.....NSAGCAEKQAKRCKEFS 1571

Scoring table: PAM 150  
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08  
 Listing first 45 summaries

Database: a-geneseq36  
 1:geneseqp

Statistics: Mean 40.973; Variance 226.720; scale 0.181

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	11180	99.9	1571	1	M42087 Human Down syndrome-ce	0.00e+00
2	11119	99.4	1910	1	M42086 Human Down syndrome-ce	0.00e+00
3	7081	18.6	465	1	M55045 Neural adhesion molecu	1.81e-145
4	740	6.6	1447	1	R68553 Deleted in colorectal	4.41e-43
5	740	6.6	1728	1	R13144 Deleted in Colorectal	4.41e-43
6	631	5.6	1018	1	R87028 Human connectin.	5.75e-35
7	629	5.6	1018	1	R63759 Human connectin (EMBL	8.08e-35
8	623	5.6	1182	1	M57900 Protein of clone CO722	2.35e-34
9	620	5.5	1257	1	M74152 Human LI cell adhesion	3.75e-34
10	606	5.4	1018	1	M06485 Rat contactin ligand f	4.08e-33
11	608	5.4	1028	1	M29667 Homo sapiens DL185.1 c	2.90e-33
12	590	5.3	761	1	R92255 Neural cell adhesion m	6.20e-32
13	588	5.3	1304	1	M59994 Human neural cell adhe	8.71e-32
14	569	5.1	1911	1	M94027 Human protein tyrosine	2.19e-30
15	569	5.1	1911	1	M27225 Human protein tyrosine	2.19e-30
16	569	5.1	1911	1	R71726 Human PTP-OB.	2.15e-30
17	549	4.9	582	1	R92256 Neural cell adhesion m	6.48e-29
18	525	4.7	1070	1	M08747 Human colon carcinoma	3.74e-27
19	514	4.6	753	1	R83927 Human T85 protein.	2.39e-26
20	486	4.3	1501	1	R72858 Rat receptor type-prot	2.64e-24
21	414	3.7	1496	1	M81030 Melanoma associated an	4.29e-19
22	394	3.5	400	1	R75203 Tyrosine phosphatase M	1.16e-17
23	353	3.2	1291	1	R75201 Tyrosine phosphatase M	9.62e-15

24	321	2.9	467	1	R84094	Nsk2 receptor with put	1.72e-12
25	319	2.9	467	1	M62575	Alternatively spliced	2.37e-12
26	319	2.9	475	1	R94982	Nsk2 extracellular dom	2.37e-12
27	322	2.9	478	1	R92718	Mouse muscle-localized	1.46e-12
28	322	2.9	860	1	R92716	Mouse muscle-localized	1.46e-12
29	319	2.9	863	1	M62569	Alternatively spliced	2.37e-12
30	319	2.9	867	1	M62583	Mouse receptor tyrosin	2.37e-12
31	322	2.9	868	1	R92717	Mouse muscle-localized	1.46e-12
32	320	2.9	868	1	M26507	Rat Dmk receptor.	2.02e-12
33	320	2.9	868	1	M26510	Rat muscle-specific k1	2.02e-12
34	326	2.9	869	1	M26506	Human Dmk receptor.	7.67e-13
35	326	2.9	869	1	M26511	Human muscle-specific	7.67e-13
36	319	2.9	871	1	R84087	Nsk2 receptor.	2.37e-12
37	319	2.9	871	1	M62568	Mouse receptor tyrosin	2.37e-12
38	319	2.9	873	1	R84092	Nsk2 receptor with alt	2.37e-12
39	319	2.9	873	1	M62573	Alternatively spliced	2.37e-12
40	319	2.9	881	1	R84091	Nsk2 receptor with alt	2.37e-12
41	319	2.9	881	1	M62572	Mouse Nsk2 (alternativ	2.37e-12
42	320	2.9	1225	1	M52289	Homo sapiens cdo tumo	2.02e-12
43	330	2.9	1242	1	M52287	Rattus norvegicus cdo	4.02e-13
44	317	2.8	863	1	R84088	Nsk2 receptor with alt	3.27e-12
45	302	2.7	1091	1	M41641	Sequence used in detec	3.63e-11

## ALIGNMENTS

RESULT 1  
 ID M42087 standard: Protein: 1571 AA.  
 AC W42087;  
 DT 28-SEP-1998 (first entry)  
 DE Human Down syndrome-cell adhesion molecule DS-CAM2.  
 KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;  
 KW signal transduction; trisomy 21; mental retardation;  
 KW holoprosencephaly; corpus callosum agenesis;  
 KW schizencephaly; diagnosis; assay; human.  
 OS Homo sapiens.  
 PN M09817795-A1.  
 PD 30-APR-1998.  
 PE 23-OCT-1997; U19547.  
 PR 25-OCT-1996; US-029322.  
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
 PI Korenberg JR.  
 DR WPI: 98-271791/24.  
 DR N-PSDB: V31988.  
 PT New isolated Down's Syndrome-cell adhesion molecule - used to  
 develop products for detection, diagnosis and therapy of  
 developmental and neurological abnormalities  
 PS Claim 2; Page 90-95; 109pp; English.  
 CC This polypeptide comprises Down syndrome-cell adhesion molecule  
 CC DS-CAM2, an extracellular soluble protein belonging to a novel  
 CC subclass of the Ig superfamily with highest homology to neural cell  
 CC adhesion molecules. Its amino acid sequence was deduced from cDNA  
 CC clones (see V31982) isolated from a trisomy 21 foetal brain library.  
 CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and  
 CC lacks the entire transmembrane domain of DS-CAM1. The invention  
 CC provides human and murine DS-CAM nucleic acid sequences (see also  
 CC V31981, V31985-87), expression vectors and host cells, transgenic  
 CC animals, antibodies, antisense oligonucleotides, and primers  
 CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated  
 CC with developmental and neurological processes. They can be used in  
 CC e.g. neural prosthetic devices used in entubulation methods of  
 CC repairing (regenerating) damaged or severed peripheral nerves, and  
 CC also in bioassays to identify agonists and antagonists. The products  
 CC can also be used in detection, diagnosis and therapy of developmental  
 CC and neurological abnormalities such as Down syndrome, mental  
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,  
 CC or schizencephaly.  
 CC Sequence 1571 AA.

Query Match 99.9%; Score 11180; DB 1; Length 1571;  
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
 Matches 1569; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

Db 1 MWLALSLFQSFANFSEEDLSLSLEYFNASLOEYVFASTGTGLVPCPAAGIPPTLRMYL 60
Oy 1 MWLALSLFQSFANFSEEDLSLSLEYFNASLOEYVFASTGTGLVPCPAAGIPPTLRMYL 60
Db 61 ANGEELYDVPGIRHVHPNCTLOIFPEPPSSFTLIHDNTYYCTAENPSGKIRSDVHIKA 120
Oy 61 ANGEELYDVPGIRHVHPNCTLOIFPEPPSSFTLIHDNTYYCTAENPSGKIRSDVHIKA 120
Db 61 ANGEELYDVPGIRHVHPNCTLOIFPEPPSSFTLIHDNTYYCTAENPSGKIRSDVHIKA 120
Oy 61 ANGEELYDVPGIRHVHPNCTLOIFPEPPSSFTLIHDNTYYCTAENPSGKIRSDVHIKA 120
Db 121 VLREPTVVEODKTRGVNAVFKCIIIPSSVAYITTVSMEDIVSLVSGSFLLTSTGA 180
Oy 121 VLREPTVVEODKTRGVNAVFKCIIIPSSVAYITTVSMEDIVSLVSGSFLLTSTGA 180
Db 121 VLREPTVVEODKTRGVNAVFKCIIIPSSVAYITTVSMEDIVSLVSGSFLLTSTGA 180
Oy 121 VLREPTVVEODKTRGVNAVFKCIIIPSSVAYITTVSMEDIVSLVSGSFLLTSTGA 180
Db 181 LYIKDVQNDGLYNTRCITRHRTGTETROSNSARLFVSPDPAASVILDEGPHRAMAQO 240
Oy 181 LYIKDVQNDGLYNTRCITRHRTGTETROSNSARLFVSPDPAASVILDEGPHRAMAQO 240
Db 181 LYIKDVQNDGLYNTRCITRHRTGTETROSNSARLFVSPDPAASVILDEGPHRAMAQO 240
Oy 181 LYIKDVQNDGLYNTRCITRHRTGTETROSNSARLFVSPDPAASVILDEGPHRAMAQO 240
Db 241 RVELPCALGHEPDYRWLKDNPLELSGRFQKIVTGLLIENIRPSDGSYCEVSNRYG 300
Oy 241 RVELPCALGHEPDYRWLKDNPLELSGRFQKIVTGLLIENIRPSDGSYCEVSNRYG 300
Db 241 RVELPCALGHEPDYRWLKDNPLELSGRFQKIVTGLLIENIRPSDGSYCEVSNRYG 300
Oy 241 RVELPCALGHEPDYRWLKDNPLELSGRFQKIVTGLLIENIRPSDGSYCEVSNRYG 300
Db 301 TAKVIGRLVVKQPLKATISPRKYSSVGSQVSLSCVCTEODELSMYRNGEILNPKNV 360
Oy 301 TAKVIGRLVVKQPLKATISPRKYSSVGSQVSLSCVCTEODELSMYRNGEILNPKNV 360
Db 301 TAKVIGRLVVKQPLKATISPRKYSSVGSQVSLSCVCTEODELSMYRNGEILNPKNV 360
Oy 301 TAKVIGRLVVKQPLKATISPRKYSSVGSQVSLSCVCTEODELSMYRNGEILNPKNV 360
Db 361 RITGINHENLIMDHVKSDBGAYOCFVRKDKLSADYVQVVLDEGTPKIIISAFSEKVVSP 420
Oy 361 RITGINHENLIMDHVKSDBGAYOCFVRKDKLSADYVQVVLDEGTPKIIISAFSEKVVSP 420
Db 361 RITGINHENLIMDHVKSDBGAYOCFVRKDKLSADYVQVVLDEGTPKIIISAFSEKVVSP 420
Oy 361 RITGINHENLIMDHVKSDBGAYOCFVRKDKLSADYVQVVLDEGTPKIIISAFSEKVVSP 420
Db 421 AEPVSLMCVWKGTPLETTITWLDDBPILKGGSHRISOMITSGANVLSYUNISSSOVRDGG 480
Oy 421 AEPVSLMCVWKGTPLETTITWLDDBPILKGGSHRISOMITSGANVLSYUNISSSOVRDGG 480
Db 421 AEPVSLMCVWKGTPLETTITWLDDBPILKGGSHRISOMITSGANVLSYUNISSSOVRDGG 480
Oy 421 AEPVSLMCVWKGTPLETTITWLDDBPILKGGSHRISOMITSGANVLSYUNISSSOVRDGG 480
Db 481 VYRGTANNAGVVLQOARINVRGPASIRPMKNITAIAGDVTYIHCRIVGPYYSLIKWKN 540
Oy 481 VYRGTANNAGVVLQOARINVRGPASIRPMKNITAIAGDVTYIHCRIVGPYYSLIKWKN 540
Db 481 VYRGTANNAGVVLQOARINVRGPASIRPMKNITAIAGDVTYIHCRIVGPYYSLIKWKN 540
Oy 481 VYRGTANNAGVVLQOARINVRGPASIRPMKNITAIAGDVTYIHCRIVGPYYSLIKWKN 540
Db 541 SNLLPFNRHOVAFENNGLTKLSDOKEVDEGEYTCNVLPOLSTOSVHYTVKVPPEIQ 600
Oy 541 SNLLPFNRHOVAFENNGLTKLSDOKEVDEGEYTCNVLPOLSTOSVHYTVKVPPEIQ 600
Db 541 SNLLPFNRHOVAFENNGLTKLSDOKEVDEGEYTCNVLPOLSTOSVHYTVKVPPEIQ 600
Oy 541 SNLLPFNRHOVAFENNGLTKLSDOKEVDEGEYTCNVLPOLSTOSVHYTVKVPPEIQ 600
Db 601 PEEFPFSSIGORVYFPCVNVSGDLPTITWOKDGRIPESLVTITDNDFTSSLSINLS 660
Oy 601 PEEFPFSSIGORVYFPCVNVSGDLPTITWOKDGRIPESLVTITDNDFTSSLSINLS 660
Db 601 PEEFPFSSIGORVYFPCVNVSGDLPTITWOKDGRIPESLVTITDNDFTSSLSINLS 660
Oy 601 PEEFPFSSIGORVYFPCVNVSGDLPTITWOKDGRIPESLVTITDNDFTSSLSINLS 660
Db 661 LMHNGVYTCIARREAAVHOSOLIVRPPKRVVOPRODDGIYGAAILNCABEYVPT 720
Oy 661 LMHNGVYTCIARREAAVHOSOLIVRPPKRVVOPRODDGIYGAAILNCABEYVPT 720
Db 661 LMHNGVYTCIARREAAVHOSOLIVRPPKRVVOPRODDGIYGAAILNCABEYVPT 720
Oy 661 LMHNGVYTCIARREAAVHOSOLIVRPPKRVVOPRODDGIYGAAILNCABEYVPT 720
Db 721 IWKFSKAGVPOFOPIALNGRIOVLSNGSLIKHVEEDSGYLCXKVSNDYGVADVSKSM 780
Oy 721 IWKFSKAGVPOFOPIALNGRIOVLSNGSLIKHVEEDSGYLCXKVSNDYGVADVSKSM 780
Db 721 IWKFSKAGVPOFOPIALNGRIOVLSNGSLIKHVEEDSGYLCXKVSNDYGVADVSKSM 780
Oy 721 IWKFSKAGVPOFOPIALNGRIOVLSNGSLIKHVEEDSGYLCXKVSNDYGVADVSKSM 780
Db 781 YLTVKIPAMITSPNTTLATOGOKKEMSTAHGEBPIYRWKEKEDRIINPEARLYVSTK 840
Oy 781 YLTVKIPAMITSPNTTLATOGOKKEMSTAHGEBPIYRWKEKEDRIINPEARLYVSTK 840
Db 781 YLTVKIPAMITSPNTTLATOGOKKEMSTAHGEBPIYRWKEKEDRIINPEARLYVSTK 840
Oy 781 YLTVKIPAMITSPNTTLATOGOKKEMSTAHGEBPIYRWKEKEDRIINPEARLYVSTK 840
Db 841 EVGEELYISTLOILPTVREDSGFFSCHAINSAGEDGIILOLTVQEPDPDEIETIKDYKART 900
Oy 841 EVGEELYISTLOILPTVREDSGFFSCHAINSAGEDGIILOLTVQEPDPDEIETIKDYKART 900
Db 841 EVGEELYISTLOILPTVREDSGFFSCHAINSAGEDGIILOLTVQEPDPDEIETIKDYKART 900
Oy 841 EVGEELYISTLOILPTVREDSGFFSCHAINSAGEDGIILOLTVQEPDPDEIETIKDYKART 900
Db 901 IYLRMTMGFDGNSPITGYDIECKNKSMDSAQRTKDVSPOLNSATIIDIHPSSTYSIRM 960
Oy 901 IYLRMTMGFDGNSPITGYDIECKNKSMDSAQRTKDVSPOLNSATIIDIHPSSTYSIRM 960
Db 901 IYLRMTMGFDGNSPITGYDIECKNKSMDSAQRTKDVSPOLNSATIIDIHPSSTYSIRM 960
Oy 901 IYLRMTMGFDGNSPITGYDIECKNKSMDSAQRTKDVSPOLNSATIIDIHPSSTYSIRM 960
Db 961 YAKNRIGKSEPSNELTITDEAAPDGPPOEVHLEPISOSIYVTKAKPKKHLONGIIRGY 1020
Oy 961 YAKNRIGKSEPSNELTITDEAAPDGPPOEVHLEPISOSIYVTKAKPKKHLONGIIRGY 1020
Db 961 YAKNRIGKSEPSNELTITDEAAPDGPPOEVHLEPISOSIYVTKAKPKKHLONGIIRGY 1020
Oy 961 YAKNRIGKSEPSNELTITDEAAPDGPPOEVHLEPISOSIYVTKAKPKKHLONGIIRGY 1020
Db 1021 QIGYREYSTGNGFQNIISVDTSGDSEVYTLDNLNKFTGYGLVQACNAGTGPSSOELI 1080
Oy 1021 QIGYREYSTGNGFQNIISVDTSGDSEVYTLDNLNKFTGYGLVQACNAGTGPSSOELI 1080
Db 1021 QIGYREYSTGNGFQNIISVDTSGDSEVYTLDNLNKFTGYGLVQACNAGTGPSSOELI 1080
Oy 1021 QIGYREYSTGNGFQNIISVDTSGDSEVYTLDNLNKFTGYGLVQACNAGTGPSSOELI 1080
Db 1081 TTLEEDVPSPYPPENVOAIIATSPESISISWSTLSKEALNGLILOGFRVIYWANLMDGELGEI 1140

```

```






Oy 1081 TTLEEDVPSPYPPENVOAIIATSPESISISWSTLSKEALNGLILOGFRVIYWANLMDGELGEI 1140
Db 1141 KNITTPQSLDGLGLEYKTYNSIQVLAFTRAGDGVSEQIFTRTKEDVGPAGVAKAAA 1200
Oy 1141 KNITTPQSLDGLGLEYKTYNSIQVLAFTRAGDGVSEQIFTRTKEDVGPAGVAKAAA 1200
Db 1201 SASWVFWSLPLKLNGLIIRKTYVFCSHYPTVYISFEFASPDSPFSIRIPNLSRNQYSW 1260
Oy 1201 SASWVFWSLPLKLNGLIIRKTYVFCSHYPTVYISFEFASPDSPFSIRIPNLSRNQYSW 1260
Db 1261 VVATVTSAGRNSSSEIITVEPLAKAPARILTFSGVTTPMKKDVLPCKAVGDPSPAVKWM 1320
Oy 1261 VVATVTSAGRNSSSEIITVEPLAKAPARILTFSGVTTPMKKDVLPCKAVGDPSPAVKWM 1320
Db 1321 KDSNGTFLSVTIDGRNISFNSGFIIRIVKAEDSGYSCSIANNMGSDEIILNLOVQVP 1380
Oy 1321 KDSNGTFLSVTIDGRNISFNSGFIIRIVKAEDSGYSCSIANNMGSDEIILNLOVQVP 1380
Db 1381 DQPLTVSKTSSITLSWLPDNGGSSIRGYTLOYSEDNSEQMSFPIPSERSYRLN 1440
Oy 1381 DQPLTVSKTSSITLSWLPDNGGSSIRGYTLOYSEDNSEQMSFPIPSERSYRLN 1440
Db 1441 LKCGTWKFTLTAQNGVGPGRISSEIIEAKTLGKEPOFSKEOELFASINTTRYRLNIGWN 1500
Oy 1441 LKCGTWKFTLTAQNGVGPGRISSEIIEAKTLGKEPOFSKEOELFASINTTRYRLNIGWN 1500
Db 1501 DGGCPIITSFLEPRFGTYTWTAAORTSLSKSYILYDLEATYELQOMVCHSAGCAEQ 1560
Oy 1501 DGGCPIITSFLEPRFGTYTWTAAORTSLSKSYILYDLEATYELQOMVCHSAGCAEQ 1560
Db 1561 AKEARCKEFS 1571
Oy 1561 AKEARCKEFS 1571

```

```

RESULT 2
ID M42086 standard: Protein; 1910 AA.
AC M42086:
DE 28-SEP-1998 (first entry)
DT Human Down syndrome-cell adhesion molecule DS-CAM1.
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis; assay; human.
OS Homo sapiens.
FH Key
FH Peptide
FT 1..23
FT /label= Sig_peptide
FT Protein
FT 24..1910
FT /label= Mal_protein
FT Domain
FT 24..887
FT /label= IG
FT /note= "Immunoglobulin type-C2 domain"
FT 888..1594
FT /label= FBN
FT Domain
FT /note= "fibronectin type III domain"
FT 1595..1616
FT /label= Transmembrane
FT 1617..1910
FT Domain
FT /label= Cytoplasmic
FT 24..126
FT /label= Ig1
FT 127..225
FT /label= Ig2
FT 226..316
FT /label= Ig3
FT 317..409
FT /label= Ig4
FT 410..506
FT /label= Ig5
FT 507..603
FT /label= Ig6

```

(TM)

```
|||||
QY 195 YRCITRRHYGTEFRQSNARLFVSDPANSAPIJLDGDFHRKAMAGQVLEJPCRALHPPE 254
Db 241 DYRWLKNMPELISGRQKTYVTGLLENIRPDSGSGYVCVSNRYGAKYIGLYKOPJ 300
QY 255 DYRWLKNMPELISGRQKTYVTGLLENIRPDSGSGYVCVSNRYGAKYIGLYKOPJ 314
Db 301 KATISPRKXSVSGVSOYSLSCSVTGTEDDELXSWYRNGEILNPKNVAITGIIHNENIMDH 360
QY 315 KATISPRKXSVSGVSOYSLSCSVTGTEDDELXSWYRNGEILNPKNVAITGIIHNENIMDH 374
Db 361 MWKSDGAGYQCFYAKDKLSADYVQVYVLEDEGTETIISAFSEKXVSPAEPYSLKCNKGP 420
QY 375 MWKSDGAGYQCFYAKDKLSADYVQVYVLEDEGTETIISAFSEKXVSPAEPYSLKCNKGP 434
Db 421 LPTITWITLDDPILKGSNHRISQMITSEGNVSYLNISSQVDRGQVYRGTANNAGVYL 480
QY 435 LPTITWITLDDPILKGSNHRISQMITSEGNVSYLNISSQVDRGQVYRGTANNAGVYL 494
Db 481 YQARINWGPASIRPMKNITAIAGROTYIHCVRIGYPYISIKWYKSNLLPFNHROYAFE 540
QY 495 YQARINWGPASIRPMKNITAIAGROTYIHCVRIGYPYISIKWYKSNLLPFNHROYAFE 554
Db 541 NNGTLKSLSDVQKEVEDEBYTCNVLPOLSTSOVHTVAVPFIOPFEPRPSIGORVF 600
QY 555 NNGTLKSLSDVQKEVEDEBYTCNVLPOLSTSOVHTVAVPFIOPFEPRPSIGORVF 614
Db 601 IPCVWVSGDLPITTTWQKDGRIPIGSLGVTIDNIDFTSSRLSNLSMHNANTCIARNE 660
QY 615 IPCVWVSGDLPITTTWQKDGRIPIGSLGVTIDNIDFTSSRLSNLSMHNANTCIARNE 674
Db 661 AAAVEHOSQILIVRPKPVVOPRDQDGIYKAVILNCSAGCYPIYVWKFSGAGYPOF 720
QY 675 AAAVEHOSQILIVRPKPVVOPRDQDGIYKAVILNCSAGCYPIYVWKFSGAGYPOF 734
Db 721 OPTLNGRIQVLSNGSLIKIHVVEDESGYIYCKVSNVGDVSKMUTYKIPAMITSY 780
QY 735 OPTLNGRIQVLSNGSLIKIHVVEDESGYIYCKVSNVGDVSKMUTYKIPAMITSY 794
Db 781 NNTLATOGOKKEMKSTAHGEXPIIVRWKEDRIINPEMARLYSTKEVGEVISTLOILP 840
QY 795 NNTLATOGOKKEMKSTAHGEXPIIVRWKEDRIINPEMARLYSTKEVGEVISTLOILP 854
Db 841 TVREDSCGFSCSHAINSGEDRGIIQLTVQEPDPPEIEIKDYKARTITLMTWGFDCNSP 900
QY 855 TVREDSCGFSCSHAINSGEDRGIIQLTVQEPDPPEIEIKDYKARTITLMTWGFDCNSP 914
Db 901 ITGYDIECKNKSQSDMSQORTKDVSPOLNSATIIDHPSSYSTRMYAKNRICKSEPSNE 960
QY 915 ITGYDIECKNKSQSDMSQORTKDVSPOLNSATIIDHPSSYSTRMYAKNRICKSEPSNE 974
Db 961 LTTIADAEAPDPGPOEYHLEPISSOSIRVTWKAPKHLQNGIIRGOYIGYREXSTGDNFQ 1020
QY 975 LTTIADAEAPDPGPOEYHLEPISSOSIRVTWKAPKHLQNGIIRGOYIGYREXSTGDNFQ 1034
Db 1021 FNITSVDTSQSGSEYVYLDNLKFTQYGLVQACNRAGTGPSSQOIIITTTLEDVPSYPEN 1080
QY 1035 FNITSVDTSQSGSEYVYLDNLKFTQYGLVQACNRAGTGPSSQOIIITTTLEDVPSYPEN 1094
Db 1081 VOATATSPESISISNSTLSKALNGILOGFRVIYWANLMDGELKINITTTOPSLELDG 1140
QY 1095 VOATATSPESISISNSTLSKALNGILOGFRVIYWANLMDGELKINITTTOPSLELDG 1154
Db 1141 LEKTYNNSIOVLATFRAGDGVRSQOIFTRKEDVPGPPAGYKAAASASVFWMLPPLK 1200
QY 1155 LEKTYNNSIOVLATFRAGDGVRSQOIFTRKEDVPGPPAGYKAAASASVFWMLPPLK 1214
Db 1201 LNGIIRKTYVCSHPYPIVISEFASPDSESYRIPNLSSRNQYVWVVAATSGRGSSE 1260
QY 1215 LNGIIRKTYVCSHPYPIVISEFASPDSESYRIPNLSSRNQYVWVVAATSGRGSSE 1274
Db 1261 IITVEPLAKAPARILTFESGVTTPMKNIDIVLPCKAVGDPSPAVVWKMDSNGTSLVITDG 1320
|||||
```

```
QY 1275 IITVEPLAKAPARILTFESGVTTPMKNIDIVLPCKAVGDPSPAVVWKMDSNGTSLVITDG 1334
Db 1321 RRSISNGSFIIRRYKABDSGYSYCIANNMNGSDEIILNLOVYPPDPQRLTYSKITSS 1380
QY 1335 RRSISNGSFIIRRYKABDSGYSYCIANNMNGSDEIILNLOVYPPDPQRLTYSKITSS 1394
Db 1381 IITLSWLPDNGGSSIRGYILOYSEDNSEQWGFPIIPSEERSYRLNLCGTWYKFTLTAQ 1440
QY 1395 IITLSWLPDNGGSSIRGYILOYSEDNSEQWGFPIIPSEERSYRLNLCGTWYKFTLTAQ 1454
Db 1441 NGVGPGRISIEIEATLTCKEPOFSKDELFASINTTRVRLNLIGMNGGCPITSTLEVR 1500
QY 1455 NGVGPGRISIEIEATLTCKEPOFSKDELFASINTTRVRLNLIGMNGGCPITSTLEVR 1514
Db 1501 PEGTWTWTTAORTSLSKSYIILYDLOEATWELQWRVNCNSAGCAEKQAN 1548
QY 1515 PEGTWTWTTAORTSLSKSYIILYDLOEATWELQWRVNCNSAGCAEKQAN 1562
|||||
```

```
RESULT 2
ENTRY 158164 #type complete
TITLE BIG-1 protein - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
```

```
ACCESSIONS 158164
REFERENCE 158164
#authors Yoshihara, Y.; Kawasaki, M.; Tanii, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K.
#journal Neuron (1994) 13:415-426
#title BIG-1: a new FGF-1/F3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity.
#cross-references PMID:94338697
#accession 158164
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-1028 ##label RES
##cross-references EMBL:U01091; NID:9563132; PIDD:AAA63607.1;
PID:9563133
```

```
GENETICS
#gene BIG-1
#classification #superfamily contactin; fibronectin type III repeat homology;
#length 1028 #molecular_weight 112788 #checksum 5866
SUMMARY
```

```
Query Match 7.2%; Score 809; DB 2; Length 1028;
Best Local Similarity 24.7%; Pred. No. 2,856-135;
Matches 250; Conservative 249; Mismatches 447; Indels 66; Gaps 54;
```

```
Db 25 GYVYKPEP-SNSIFP--V-GSEWKKIILNCEARGNPSPHYRWOLNGSDIDTSLDHRK-L 79
QY 306 GRLYXKQPLKATISPRKXSVSGVSOYSLSCSVTGTEDDELXSWYRNGEILNPKNVAITGI 365
Db 80 NCGNLIIVIPNRMNMTGSGYOCFA-TNSIGTIVSREAKLOFAYLENFKSRMSRVSVREGQ 138
QY 366 NHENLIMDHMKVS-DGAGYQCFVRKDKLSADYVQVYVLEDEG-TKIIISAFSEKY-VSPA 422
Db 139 GYVLLCGPPHSGELSYAVFENEYPSFEVEDSRREVSQETHLYIAKVPDVGNYTCVV 198
QY 423 PYSLMCNVWG-TPLPTIWLTLDDP-ILKGSNR-ISO---M-ITS-E-GNVSYLNIS 472
Db 199 TSTYVNAKYLGSPTLVLRSDCYMGEYEPKTELOFPETLPAKQ-SYV-K---LECFAL 252
QY 473 SSQVDRGQVY-RCIA-N-NSAGVY-LYQARINWGPASIRPMKNITAIAGROTYIHCVR 528
Db 253 GNPVQIWN-RSDDMPEPTK-IKLKFNGLYEIPNFQOE-DTGSYEC-IENSRGKNVA 308
QY 529 GYPIYSIKKYKSNLLPFRNHRVAFEN-NGTLKLSQVQKEVEDEBYTCNVLPOLSTSQ 587
Db 309 KGRLTYYAKPYWVOLLKDVETAVEDSLYMEGRA-SGRKPSYRWLKNQDAL--VLEERIQ 365
QY 588 SVHVTYKVPF-IQFEPRPSIGORVFIQVWVSGDLPITITWQKDGRIPIGSLGVTID 646
|||||
```

\*\*\*\*\*  
 N O T E S  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution Rights by Oxford Molecular Ltd

mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Apr 25 13:32:01 2000: Maspar time 48.10 Seconds

Tabular output not generated.

Title: >US-08-956-991-11

Description: (1-1571) from US08956991A.pep

Perfect Score: 11189

Sequence: 1 MMLALSLQSFANVSEDL.....NSAGCAKQAKAKRCKFS 1571

Scoring table: PAM 150

Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 56.705; Variance 92.824; scale 0.611

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	744	6.6	1447	1 DCC_MOUSE	TUMOR SUPPRESSOR PROTE	1.11e-141
2	740	6.6	1447	1 DCC_HUMAN	TUMOR SUPPRESSOR PROTE	1.06e-140
3	720	6.4	1040	1 AXOI_HUMAN	AXONIN-1 PRECURSOR (AX	8.57e-136
4	703	6.3	1036	1 AXOI_CHICK	AXONIN-1 PRECURSOR (AX	1.24e-131
5	701	6.3	1040	1 AXOI_RAT	AXONIN-1 PRECURSOR (AX	3.84e-131
6	677	6.1	1284	1 NRCA_CHICK	NC-CAM RELATED CELL AD	2.78e-125
7	663	5.9	1010	1 CAMT_CHICK	CONTACTIN PRECURSOR (N	7.14e-122
8	637	5.7	1259	1 CAMT_RAT	NEURAL CELL ADHESION M	1.47e-115
9	637	5.7	1260	1 CAMT_MOUSE	NEURAL CELL ADHESION M	4.19e-114
10	631	5.6	1018	1 CONT_HUMAN	CONTACTIN PRECURSOR (G	4.19e-114
11	631	5.6	1020	1 CONT_MOUSE	CONTACTIN PRECURSOR (N	1.93e-111
12	620	5.5	1257	1 CAML_HUMAN	NEURAL CELL ADHESION M	8.76e-109
13	609	5.4	2029	1 LAR_DROME	PROTEIN-TYROSINE PHOSP	3.94e-106
14	598	5.3	1897	1 PTFP_HUMAN	PTPF_HUMAN	2.11e-99
15	570	5.1	837	1 NCM2_MOUSE	NEURAL CELL ADHESION M	2.32e-100
16	574	5.1	1912	1 PTPD_HUMAN	PROTEIN-TYROSINE PHOSP	1.75e-97
17	562	5.0	837	1 NCM2_HUMAN	NEURAL CELL ADHESION M	1.75e-97
18	558	5.0	1091	1 NCAL_CHICK	NEUROGLIN PRECURSOR.	1.78e-96
19	562	5.0	1239	1 NRGE_DROME	NEUROGLIN PRECURSOR.	6.03e-93
20	543	4.9	848	1 NCAL_HUMAN	NEURAL CELL ADHESION M	2.01e-88
21	524	4.7	1088	1 NCAL_XENTIA	NEURAL CELL ADHESION M	3.48e-88
22	523	4.7	1092	1 NCAL_XENTIA	NEURAL CELL ADHESION M	4.37e-90
23	531	4.7	1266	1 NCAL_CHICK	NEURONAL-GLIAL CELL AD	

24	509	4.5	761	1 NCAL_HUMAN	NEURAL CELL ADHESION M	7.26e-85
25	498	4.5	853	1 NCAL_BOVIN	NEURAL CELL ADHESION M	2.89e-82
26	492	4.4	858	1 NCAL_RAT	NEURAL CELL ADHESION M	7.52e-81
27	496	4.4	1115	1 NCAL_MOUSE	NEURAL CELL ADHESION M	8.57e-82
28	478	4.3	725	1 NCAL_MOUSE	NEURAL CELL ADHESION M	1.48e-77
29	401	3.6	898	1 FAS2_SCHAM	FASCICLIN II PRECURSOR	1.11e-59
30	402	3.6	3707	1 PGBM_MOUSE	BASEMENT MEMBRANE-SPEC	6.53e-60
31	378	3.4	811	1 FS22_DROME	FASCICLIN II, PHOSPHAT	1.94e-54
32	378	3.4	811	1 FS21_DROME	FASCICLIN II, MEMBRANE	1.94e-54
33	364	3.3	4933	1 PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	2.84e-51
34	313	2.8	2477	1 FINC_RAT	FIBRONECTIN PRECURSOR	6.17e-40
35	308	2.8	2477	1 FINC_MOUSE	FIBRONECTIN PRECURSOR	7.64e-39
36	313	2.8	2481	1 UN52_CAEEL	BASEMENT MEMBRANE PROT	6.17e-40
37	287	2.6	345	1 OPCM_BOVIN	OPC10D BINDING PROTEIN	2.69e-34
38	286	2.6	1131	1 MYEP_CHICK	MYOSIN-BINDING PROTEIN	4.41e-34
39	294	2.6	2265	1 FINC_BOVIN	FIBRONECTIN (FN).	8.36e-36
40	284	2.5	345	1 OPCM_HUMAN	OPC10D BINDING PROTEIN	1.18e-33
41	279	2.5	345	1 OPCM_RAT	OPC10D BINDING PROTEIN	1.39e-32
42	264	2.4	337	1 G55A_CHICK	NEURITE INHIBITOR GP35	2.08e-29
43	267	2.4	1142	1 MYEP_HUMAN	MYOSIN-BINDING PROTEIN	4.88e-30
44	262	2.3	333	1 AMAL_DROME	AMALGAM PROTEIN PRECUR	5.51e-29
45	262	2.3	2386	1 FINC_HUMAN	FIBRONECTIN PRECURSOR	

# ALIGNMENTS

RESULT 1 STANDARD: PRT: 1447 AA.

AC p70211:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.  
 GN DCC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BRAIN:  
 RX MEDLINE; 96112625.  
 RA COOPER H.M., ARMES P., BRITTO J., GAD J., WILKS A.F.;  
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer  
 gene (DCC) and its expression in the developing mouse embryo."  
 RL Oncogene 11:2243-2254(1995).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN-BALB/C; TISSUE-BRAIN:  
 RA COOPER H.M.:  
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM  
 CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD  
 CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE  
 CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN  
 CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.  
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS  
 CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION  
 CC AND REMAIN AT THIS LEVEL IN THE ADULT.  
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL: X85788; CAAS9786.1;
DR HSP: P56276; 1TLK-
DR MGD: MG1.94869; DCC.
DR PFM: PF00047; 1g; 4.
DR PFM: PF00047; 1g; 4.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1447
FT CHAIN 85 1447
FT CHAIN 26 1097
FT TRANSMEM 1098 1122
FT DOMAIN 1123 1447
FT DOMAIN 54 124
FT DOMAIN 154 219
FT DOMAIN 254 317
FT DOMAIN 345 407
FT DOMAIN 426 522
FT DOMAIN 525 618
FT DOMAIN 619 716
FT DOMAIN 722 816
FT DOMAIN 840 940
FT DOMAIN 941 1042
FT DOMAIN 61 117
FT DISULFID 161 212
FT DISULFID 261 310
FT DISULFID 352 400
FT CARBOHYD 60 60
FT CARBOHYD 94 94
FT CARBOHYD 298 299
FT CARBOHYD 318 318
FT CARBOHYD 478 478
FT CARBOHYD 628 628
FT CARBOHYD 702 702
FT VARSPLIC 819 838
FT SEQUENCE 1447 AA; 158298 MW; F5FB79BA CRC32;

Query Match 6.68; Score 744; DB 1; Length 1447;
Best Local Similarity 29.68; Pred. No. 1,11e-141;
Matches 228; Conservative 178; Mismatches 293; Indels 72; Gaps 55;

Db 72 IKKKKGLLALGMDRRKQDLPGNSLLIONILHSRHKDEGLYOCASLADSGSIISRT 131
Oy 535 IKMYKNS-NL-LFNNHROVAFENNGTL-K-LSDVQKEVDEGGYTCNV-LVQPLSTSQS 588
Db 132 AKYTVAGPLRLFSQTESITAF-MGDTVLKCEYIGEPMP-TIMQNNODL-NPLRGDSR 188
Oy 589 VHTVQVVP-FEIQPFE-FPRFSIGORVFTPCVVSGDLPITIMQKGRIPGSLGVTID 646
Db 189 VVVLPSGALDISLQPDGSGVYRCSARNPASIRTGNAEVRILSDPGLHROLTYLQRPEN 248
Oy 647 NIDETS-SLRTISNLNHNNGNYTCLIANEAA-VEHOSOL-IVRV--PK--FVVG-PRD 698
Db 249 VIAIEGDAVLECCVSGYPPPSFTW-L-RGEVYQLR-S-K-KYSLGGSNLLISNVTD 302
Oy 699 QDGIYGAAILNCAEGEYPTIWKFKSGAGVQFQPIALNGRIOVLSNGSILLIKHAYE 758
Db 303 DDGCTYTCVVTYK-NENISASALTYLVPPWFLNHPNLTAYSMDEFECAVSG-KPV 360
Oy 759 EDSGYLCKVSNFVAGVDSKMYLTWKIPMTITSYPTTLATGOKKEMSTHAGEKPT- 817
Db 361 TVNMKNGDVVIPS-D-YF-QI--VGG--SNLRILGVKSDGEFYOCVANEAGNAQSS 412
Oy 818 IVNKEKEIRIINPEMARIYLVSTKEVGEVISTLOILPTVEDSDSFCCHAINSAGEDRG 877
Db 413 AOLIVPRAPISSTILPSAPDVLPVLVSSFRVLSMRPAKAGNIQTTFVFSRGDN 472
Oy 878 IQLTQEPDPPE-I--EI-KDV-----KARTITLRWTMGFDGSPITGVDIECKNKSDS 928
Db 473 REBALNTQPGS-LQ-LTVGNLKEPAMTFRVAVYNEGSPSSQPIKVAIOTPELIQVGP 530
Oy 929 WDSQORTKDVSPOLNSATITIDHSSTYSIMYAKNRIGKSEPSNELTITAD-EAARDGP 987

```

```

Db 531 VENVLAIVSYSPSILITWEPRA-Y-ANGPVQYRLCFTEVSTGKE-Q-NI-EVD--GLS- 582
Oy 988 POEVLHEPISQSIRVITWAKPKHLLQNGITRGYQIGREYTGNGFQFNITSVTSDSD 1047
Db 583 -YKLEIGKFTETLRFALNRYGPGVSTDDITVTLSDVSAPPONISLEVNSRIKV 641
Oy 1048 VYTLNLNFTYGLVQVACNAGCGSPSQEITITLEDVSPYENVOALATSPESISI 1107
Db 642 SWLPPSGTNGCFITGYKIRHKKTRRGDM-ETLE-PNNLWTL-FTLEKSGOVSFOVSA 698
Oy 1108 SWTLSEKALNLIQGFVRYVWNLMDLGEIKNTTQPSLELDGLEYKTNISIOVLA 1167
Db 699 MYNCGTPSNMVTAEETPNDDESQVDPDPSLHVHPQTCII-MSWTPPLNPIVYRG 757
Oy 1168 FTRAGGVASE--QIFTRKE-D---VPGPPAGVAAAASAMVSVSLPPLKNGIIRK 1221
Db 758 YIIGVGSGPYAEYRV-DSK-QRY-YSTIERLESSHYISLKAFFNNGEG 805
Oy 1222 YTV-F-CSHPYPTVISEFASPSFSYRIPNLNRNOYVWVAVTSAGRG 1270

RESULT 2
ID DCC_HUMAN STANDARD; PRT: 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9501532.
RA HEDRICK L., CHO K.R., FEARON E.R., WU T.-C., KINZLER K.W.,
RT "The DCC gene product in cellular differentiation and colorectal
tumorigenesis."
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE: 90100559.
RA FEARON E.R., CHO K.R., NIGRO J.M., KERN S.E., SIMONS J.W.,
RA RUPPERT J.M., HAMILTON S.R., PREISINGER A.C., THOMAS G., KINZLER K.W.,
RT "Identification of a chromosome 18q gene that is altered in
colorectal cancers."
RL Science 247:49-56(1990).
RN [3]
RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBELED EXONS).
RX MEDLINE: 91121517.
RA NIGRO J.M., CHO K.R., FEARON E.R., KERN S.E., RUPPERT J.M.,
RA OLINER J.D., KINZLER K.W., VOGELSTEIN B.,
RT "Scrambled exons."
RL Cell 64:607-613(1991).
RN [4]
RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX MEDLINE: 94245241.
RA CHO K.R., OLINER J.D., SIMONS J.W., HEDRICK L., FEARON E.R.,
RA PREISINGER A.C., HEDGE P., SILVERMAN G.A., VOGELSTEIN B.,
RT "The DCC gene: structural analysis and mutations in colorectal
carcinomas."
RL Genomics 19:525-531(1994).
RN [5]
RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX MEDLINE: 94243823.
RA MIYAKE S., NAGAI K., YOSHINO K., OTO M., ENDO M., YUASA Y.,
RT "Point mutation and allelic deletion of tumor suppressor gene DCC in
human esophageal squamous cell carcinomas and their relation to
Cancer Res. 54:3007-3010(1994).
CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.

```



```

OY 241 RVELPCALGHPEDDYBMLKDNMPLSLGSRFOKTVTGLLIENIRPSPSGSYCEVSNRYG 300
DB 301 TAKYIGLYKOPKATISPRKXVSSVGSVLSGCVTGEDEDELSRYNGELINGKNV 360
OY 301 TAKYIGLYKOPKATISPRKXVSSVGSVLSGCVTGEDEDELSRYNGELINGKNV 360
DB 361 RITGINENLIMDMHVKSDGAYOCFVRKDKLSADYVQVVLLEDGTPKIIISAESEKVVSP 420
OY 361 RITGINENLIMDMHVKSDGAYOCFVRKDKLSADYVQVVLLEDGTPKIIISAESEKVVSP 420
DB 421 AEPVSLMCNVKGTPLPTITWTLDDPILKGGSHRISQMITSEGNVSYLNISSQVARDG 480
OY 421 AEPVSLMCNVKGTPLPTITWTLDDPILKGGSHRISQMITSEGNVSYLNISSQVARDG 480
DB 481 VYRGTANNSAGVVLVQARINVRGPASIRPMKNITAIAGROTTHICRYIGYPIYSIKMYKN 540
OY 481 VYRGTANNSAGVVLVQARINVRGPASIRPMKNITAIAGROTTHICRYIGYPIYSIKMYKN 540
DB 541 SNLJPFNRHQAFAFENNGTLKLSDVQKEVEGEYTCNVLPQSTSSVHVYKVPFIQ 600
OY 541 SNLJPFNRHQAFAFENNGTLKLSDVQKEVEGEYTCNVLPQSTSSVHVYKVPFIQ 600
DB 601 PPEPRFSGRVPFPCVVSQGLPITITWQKDRPIPGSLGVTIDNIDFTSSLRISNLS 660
OY 601 PPEPRFSGRVPFPCVVSQGLPITITWQKDRPIPGSLGVTIDNIDFTSSLRISNLS 660
DB 661 LMHNGNYTCIARNEAAVEHOSQLIVRPPKVVQOPRQDGIYKAVILNCSAGYEVPT 720
OY 661 LMHNGNYTCIARNEAAVEHOSQLIVRPPKVVQOPRQDGIYKAVILNCSAGYEVPT 720
DB 721 IWKFSKAGVPOFOPIALNRIQVLSNGLIKHVEDESGYLLCKVSNVGVADVSKSM 780
OY 721 IWKFSKAGVPOFOPIALNRIQVLSNGLIKHVEDESGYLLCKVSNVGVADVSKSM 780
DB 781 YLTVAIPAMITSYPTTLATOGOKKEMKSTAGKEPIIVRREKEDRIINPMAYILVSTK 840
OY 781 YLTVAIPAMITSYPTTLATOGOKKEMKSTAGKEPIIVRREKEDRIINPMAYILVSTK 840
DB 841 EVGEEVISTLOILPTVRDESGFFSCHAINSXYGDRGIIQLTVQEBPDPPELEIKVKART 900
OY 841 EVGEEVISTLOILPTVRDESGFFSCHAINSXYGDRGIIQLTVQEBPDPPELEIKVKART 900
DB 901 IYLRMTMGFDGNSPIITGYDIECKNKSQRTKDVSPOLNSATIIDIHPSYSIRSM 960
OY 901 IYLRMTMGFDGNSPIITGYDIECKNKSQRTKDVSPOLNSATIIDIHPSYSIRSM 960
DB 961 YAKNRIGKSEPNELTITADEAAPDGPQEVHLEPISQSIRVYWKAKRKHLDONGIIRGY 1020
OY 961 YAKNRIGKSEPNELTITADEAAPDGPQEVHLEPISQSIRVYWKAKRKHLDONGIIRGY 1020
DB 1021 QIGYREYSTGNGFOENIISVDTSGSEYVTLTDNLKFTQYGLVQACNRAGTGPSSQCI 1080
OY 1021 QIGYREYSTGNGFOENIISVDTSGSEYVTLTDNLKFTQYGLVQACNRAGTGPSSQCI 1080
DB 1081 TTTLEDEVPYPPENVOALATSPESISISWSTLSKALNGILOGFVITWAMIMDELECI 1140
OY 1081 TTTLEDEVPYPPENVOALATSPESISISWSTLSKALNGILOGFVITWAMIMDELECI 1140
DB 1141 KNTTTPSLLELDGLEYKTYNSIOVLATFRAGDGVASEQIFFTKEDVPGPAGYKAAA 1200
OY 1141 KNTTTPSLLELDGLEYKTYNSIOVLATFRAGDGVASEQIFFTKEDVPGPAGYKAAA 1200
DB 1201 SASMNVSVMLPLKNGIIRKTYTFCSHRYPTVISEFEASPSFSYRIPLNLSRNOIYSW 1260
OY 1201 SASMNVSVMLPLKNGIIRKTYTFCSHRYPTVISEFEASPSFSYRIPLNLSRNOIYSW 1260
DB 1261 VVAVTISAGGNSSEITVEPLAKAPARILFTSGIVTTPMKMDIVLPCKAVGDPSPAVKWM 1320
OY 1261 VVAVTISAGGNSSEITVEPLAKAPARILFTSGIVTTPMKMDIVLPCKAVGDPSPAVKWM 1320
DB 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNNGSDEIIILNLOVQVPP 1380
OY 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNNGSDEIIILNLOVQVPP 1380

```

```

OY 1321 KDSNGTSPSLVTTIDGRSISNGSFIIIRYKAEDSGYSCSIANNNGSDEIIILNLOVQVPP 1380
DB 1381 DQPLRVSKTSSITLMLPDNGSSIRGYILQYSEDNSEQSGSPISPSERSYLEN 1440
OY 1381 DQPLRVSKTSSITLMLPDNGSSIRGYILQYSEDNSEQSGSPISPSERSYLEN 1440
DB 1441 LKCGTWYKFTTLAONGVPGRISEIIEKTLGKRPQSKDELFASINTRVRLNLGN 1500
OY 1441 LKCGTWYKFTTLAONGVPGRISEIIEKTLGKRPQSKDELFASINTRVRLNLGN 1500
DB 1501 DGGCPTSTLETPRFGTTWTWTAORTSLSKSYLLDQEAATWELQMRVNSAGCAEQ 1560
OY 1501 DGGCPTSTLETPRFGTTWTWTAORTSLSKSYLLDQEAATWELQMRVNSAGCAEQ 1560
DB 1561 AKEARCKEFS 1571
OY 1561 AKEARCKEFS 1571

RESULT 2
ID 060468 PRELIMINARY: PRT: 1896 AA.
AC 060468;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE (FRAGMENT).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF023449; AAC17966.1;
DR PFAM: PF00041; fn3: 6.
DR PFAM: PF00047; fn3: 8.
FT NON_TER
SO SEQUENCE 1896 AA: 209785 MW: 0DESEQCE CRC32:

Query Match 98.4%; Score 11009; DB 4; Length 1896;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1547; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 1 VSEDLHSSLIYFNASLOEVYFASTTGTLVPCPACGIPPVTLRWTLATGELIYDVPGIRH 60
OY 15 VSEDLHSSLIYFNASLOEVYFASTTGTLVPCPACGIPPVTLRWTLATGELIYDVPGIRH 74
DB 61 VHPNGTLOIFPPPPSFSSTLIHUNTYCTAENPSGKIRSQDVHKAVALREPTYVVEDOK 120
OY 75 VHPNGTLOIFPPPPSFSSTLIHUNTYCTAENPSGKIRSQDVHKAVALREPTYVVEDOK 134
DB 121 TMRGNVAFKCIIPSSVEAYITVSWEKDTVLSVSGSRFLITSTGALYIKDVQNEDEGLN 180
OY 135 TMRGNVAFKCIIPSSVEAYITVSWEKDTVLSVSGSRFLITSTGALYIKDVQNEDEGLN 194
DB 181 YRCTRHRYTGETRQSNARSARLFVSDPANSAPSLDGFDRHKKAMAGRGVLEPCKALGHPER 240
OY 195 YRCTRHRYTGETRQSNARSARLFVSDPANSAPSLDGFDRHKKAMAGRGVLEPCKALGHPER 254
DB 241 DYRWLKDMPLELSGRFOKTVTGLIENIRPSPSGSYCEVSNRYGTAKVIGRLYKQPL 300
OY 255 DYRWLKDMPLELSGRFOKTVTGLIENIRPSPSGSYCEVSNRYGTAKVIGRLYKQPL 314
DB 301 KATISPRKXVSSVGSQVLSGCVTGEDEDELSRYNGELINRPGKNRITGINENLIMDH 360
OY 315 KATISPRKXVSSVGSQVLSGCVTGEDEDELSRYNGELINRPGKNRITGINENLIMDH 374
DB 361 MYKSDGAYOCFVRKDKLSADYVQVVLLEDGTPKIIISAESEKVVSPAEPVSLMCNVKGT 420
OY 375 MYKSDGAYOCFVRKDKLSADYVQVVLLEDGTPKIIISAESEKVVSPAEPVSLMCNVKGT 434

```



OM of: US-08-956-991-11 to: GenEmbl: \* out\_format : pfs

Date: Apr 26, 2000 7:24 AM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US08956991/runatc\_25042000\_121637\_9559/app-query.fasta.1  
-DB=GenEmbl -QEMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.000 -LOOPEL=0.000 -LOOPEXT=0.000 -FGAPOP=6.000  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500 -TRANS=human40.cdi  
-DELXT=7.000 -START=1 -MATRIX=blsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=1000000 -USER=US08956991  
-NCPU=6 -ICPU=3 -NO\_XLPHY -WAIT -THREADS=1

#### Search information block:

Query: US-08-956-991-11

Query Length: 1571

Database: GenEmbl: \*

Database sequences: 821193

Search length: -1518192014

Search time (sec): 3401.920000

#### score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr3:AF023450	8223.00	9658.14	0.0	6413	AF023450 Homo sapiens CHD2-52
gb_pr3:AF023449	8101.00	9554.38	0.0	6110	AF023449 Homo sapiens CHD2-42
gb_pr2:AB032958	4570.00	5378.18	1.5e-291	6025	AB032958 Homo sapiens mRNA for
gb_hlg7:AC017967	1236.00	1416.45	7.1e-71	31646	AC017967 Drosophila melanogaster
gb_hlg7:AC014917	903.00	1027.16	3.4e-49	20940	AC014917 Drosophila melanogaster
gb_hlg7:AC014917	878.00	1006.45	4.9e-48	9412	U88578 Drosophila melanogaster
gb_hlg7:AC010115	876.50	979.60	1.5e-45	90563	AC010115 Drosophila melanogaster
gb_pr3:AF042091	853.00	948.59	8.1e-45	121019	AF042091 Homo sapiens chromo
gb_pr3:AF042090	853.00	945.82	1.2e-44	153407	AF042090 Homo sapiens chromo
gb_hlg7:AC017825	829.50	933.46	5.6e-44	38583	AC017825 Drosophila melanogaster
gb_pr1:HSITITINN2	815.50	922.21	4.4e-43	23887	X90569 H. sapiens mRNA for tit
gb_pr1:RNU68726	808.50	931.42	7.3e-44	4924	U68726 Rattus norvegicus Deleted
gb_pr1:RNU68725	807.50	931.19	7.5e-44	4517	U68788 M. musculus mRNA for neural c
gb_pr1:RNU68725	798.00	906.23	1.9e-42	15597	AF074901 Caenorhabditis elegans
gb_pr1:RNU68725	794.00	915.00	6.0e-43	4608	X76132 H. sapiens DCC mRNA, 8/1
gb_pr1:RNU68725	792.50	909.3	1.8e-41	5093	X59149 Rat mRNA for neural cell
gb_pr1:RNU68725	764.50	879.56	5.7e-41	4846	U10986 Xenopus laevis deleted
gb_pr1:RNU68725	762.50	877.67	7.2e-41	4641	U68726 Rattus norvegicus neog
gb_pr1:RNU68725	758.50	875.20	9.9e-41	378	X12875 Mouse mRNA for neural c
gb_pr1:RNU68725	750.50	863.53	4.4e-40	4619	X14877 Gallus gallus mRNA for
gb_pr1:RNU68725	749.50	861.49	5.8e-40	4991	Y00813 Chicken mRNA for contac
gb_pr1:RNU68725	748.50	863.40	4.5e-40	3774	AR036493 Sequence 1 from paten
gb_pr1:RNU68725	748.50	863.07	4.7e-40	3888	M74387 Human cell adhesion mol
gb_pr1:RNU68725	744.50	863.39	5.8e-40	4523	M76640 Homo sapiens LI cell ad
gb_pr1:RNU68725	744.50	863.14	1.7e-39	6218	AB002341 Human mRNA for KIAA03
gb_pr1:RNU68725	742.00	861.72	2.0e-39	5414	Y09535 M. musculus mRNA for ne
gb_pr1:RNU68725	741.50	854.13	1.5e-39	4128	U11031 Rattus norvegicus neur
gb_pr1:RNU68725	740.50	853.23	1.7e-39	4023	AF129167 Chlorocephus aethiops
gb_pr1:RNU68725	734.50	844.57	5.0e-39	4631	U72391 Human neogenin mRNA, c
gb_pr1:RNU68725	734.50	840.72	8.3e-39	5297	U61262 Human neogenin mRNA, c
gb_pr1:RNU68725	732.50	843.28	5.9e-39	3218	D86505 Xenopus laevis mRNA for
gb_pr1:RNU68725	730.00	843.28	6.8e-39	3341	D87212 Rattus norvegicus mRNA
gb_pr1:RNU68725	729.50	843.07	6.8e-39	3801	U55211 Carassius auratus LI-1
gb_pr1:RNU68725	728.00	839.27	1.0e-38	3910	AB015205 Xenopus laevis mRNA f
gb_pr1:RNU68725	727.50	840.15	8.7e-39	3210	AB013802 Homo sapiens mRNA for
gb_pr1:RNU68725	727.50	838.33	1.1e-38	3900	AB010107 Homo sapiens mRNA for
gb_pr1:RNU68725	725.50	835.83	1.5e-38	3774	X59847 H. sapiens mRNA for neur
gb_pr1:RNU68725	725.50	835.01	1.7e-38	3412	D32135 Bovine mRNA for F3/F11/
gb_pr1:RNU68725	722.50	833.76	2.0e-38	3442	L01991 Mus musculus TAF6-1-11x
gb_pr1:RNU68725	719.50	830.10	3.2e-38	3448	Y17793 Mus musculus mRNA for D
gb_pr1:RNU68725	719.50	829.36	3.5e-38	3314	U07819 Human contactin 1 precu
gb_pr1:RNU68725	718.50	829.29	3.6e-38	3335	U07820 Human contactin 2 precu

gb\_pr1:RNU68725 + 718.50 827.72 4.4e-38 3843 X14943 Mouse mRNA for neuro  
gb\_pr1:RNU68725 + 717.50 828.02 4.2e-38 3360 AR001474 Sequence 5 from pa  
gb\_pr1:RNU68725 + 717.50 828.02 4.2e-38 3360 I74665 Sequence 5 from pate

seq\_name: gb\_pr3:AF023450

seq documentation block: 6413 bp PRI 01-JUN-1998

LOCUS AF023450 Homo sapiens CHD2-52 Down syndrome cell adhesion molecule (DSCAM)

DEFINITION mRNA, complete cds.

ACCESSION AF023450

VERSION AF023450.1 GI:3169767

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 6413)

Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,

Lyons, C.E. and Korenberg, J.R.

DSCAM: a Novel Member of the Immunoglobulin Superfamily Maps in a

Down Syndrome Region and is Involved in the Development of the

Nervous System

Unpublished

2 (bases 1 to 6413)

Yamakawa, K., Huo, Y.-K., Haendel, M.A., Hubert, R., Chen, X.-N.,

Lyons, C.E. and Korenberg, J.R.

Direct Submission

Submitted (08-SEP-1997) Medical Genetics, Cedars-Sinai Research

Institute, 110 George Burns Road, Davis Building, Suite 2005, Los

Angeles, CA 90048-1869, USA

Location/Qualifiers

1. 6413

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21 (trisomy 21)"

/map="21q22, between HMG14 and MX1"

/tissue\_type="brain"

/dev\_stage="14 weeks, fetal"

/clone="CHD2-52"

/note="derived from alternately-spliced mRNA"

1. 6413

/gene="DSCAM"

453. 5168

/gene="DSCAM"

/note="member of immunoglobulin superfamily. involved in

nervous system development"

/codon\_start=1

/product="Down syndrome cell adhesion molecule"

/protein\_id="AAC17967.1"

/db\_xref="GI:3169768"

/translation="MMILALSLSFSSNFSEDLHSSLYEFVNASLQEVFASTGTLY

PCPAGAPVPTLVATLGEITVYPCGRHVPNGTQAIPEPSSSEVETITVYVM

TAENPSKIRSDVHKAVEREPTVREDKTRNGVNAFKCIPSSVEYITVYVM

EKDVSILVSGRFLITSGALYIKDYNEDLYRRTIRRTYGTGRQNSARLEVS

EPDVAAPSLDGFDRKAMAGORVELCKAPPEPRYRKMDKPELSEGRKVTY

GLIENIRPSDGSVCEVNSRGTAIVTGLYKOPDKATISPRKXSVGSVSLIS

CSVCTEDDELRSKNGELINPGRNRTTGTNNHNLIMDHVKSDDGAGVCFVKDL

SHQDYVOYLEDGTPKTIISAESEKVSAPSVSLMCVWKGKPLPTITMTDDPDLIK

GSHSRISQITSEGVNVSSTLNTSSVQDGVYRGTANNASVULYQRIYRGPASIR

PMKNTAIAGMDITHCVRVIGYPPYSIKMYNSMLPNNHQAVERENGTUKLSQVK

EYDEGEYCNLVQPLSTQSVHATVYVPEFIQPEFRPSIQARVFIQCVVSGDL

PIITWQDGRPIRGSLSTVDITNDITFSLAISLSLHNGNYCIARNEAAVEHS

OLIVRPPKFPVQPDQDGYTKAVILNCSAEGVDPVTVKFSKGAQVQFQPIALN

ATQGVKMSCTAHKEKPIIVRWEEDKIINPENARLVSTKEVEEYISTQLTPLY

REDSGFSCHAINSYGEDRGITLVQPPPEIKDVAARTILTKAKNIGKSS

ITGYDIECKNSDWSAQRTKDVSPQNSATIIDIHSSVYSIRMAKNIGKSS

NELITTADEAPDPGPEVHELEPSSGIRVYMAKPKKHLONGIRISQIGREYSD

GNFOFNITSVTSDGSEVYTLNDINKFQVGLVVOACBAGTGSSEIITITLEDV

STPPENVAIATSPESISISSTSKALNLIIDGFRVIVANIMDELDIEIKNITTT

OPSLDPLKLTNTNISIVLAFTRAGDVSRESEQIFTRKEDVPPPPAGVAAASAM

VFVSWLPPLKLTNGIIRKTYVCSHPYVPIVISEFASPDSEFSYIRNLRSNKQISWVY

AVTSAGRNSSEITVEPLAKAPARILFESGVTTPMKMDIVLPCKAVGDPSPAVKRM  
KSNQSTPSLVTIDGRRSIFJNSGFIIRTKAEDSGYVACINNNMGSDSEITLANTOVY  
PPDQRLTYSKTITSSITISLWLPDNGSSLRGYLQI SEDNSEMGSPISPSERSY  
RLENLKCGTWKFTLQONGVGRISIEIKETGKRPQSKDELFASINTTRVRL  
NLIGNNDGCCPISTSEFLXRPPIGTWTMTAORTLSKSYILYDLOEATWYELQWRVCN  
SAGCAEKQKEAARCKEFS"

BASE COUNT 1633 a 1779 c 1709 g 1292 t  
ORIGIN

## alignment\_scores:

Quality: 8223.00 Length: 1571  
Ratio: 5.234 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-956-991-11 x AF023450 ..

Align seg 1/1 to: AF023450 from: 1 to: 6413

1 MetTPIleLeuAlaLeuSerLeuPheGlnSerPheAlaAsnValPheSe 17  
|||||  
453 ATGTGATACAGCTCTCTCTCTCCAGAGCTTCGGAATGTTTTCAG 502  
17 TGIAspLeuHisSerSerLeuTyrPheValAsnAlaSerLeuGlnLuv 34  
|||||  
503 TGAAGACCTACACTCCAGCCCTCTCTTTCATGCATCTCTGCAAGAG 552  
34 a1Val1PheAlaSerThrThrGlyThrLeuValProCysProAlaAlaGly 50  
|||||  
553 TAGGTITGCCACACACGCGGACTGTGTCCTGCCCGCAGCAGCG 602  
51 lleProValThrLeuArgTrpTyrLeuAlaThrGlyGlnGlyLeu 67  
|||||  
603 ATCCCTCTGTGACTGTACATGGTACCTACGCCAGGCGAGAGATCTA 652  
67 rAspValProGlyIleArgHisValHisProAsnGlyThrLeuGlnIleP 84  
|||||  
653 CGATGTCCCGGATCCGCCAGCTCCACCCCAACGGCAGCTCTCAATT 702  
84 heProPheProProSerSerPheSerThrLeuIleHisAspAsnThrTyr 100  
|||||  
703 TCCCTTCCCTCTTCAAGCTTCAAGCTTAATCATGATTAATACTAT 752  
101 TyrCysThrAlaGluAsnProSerGlyLysIleArgSerGlnAspValHi 117  
|||||  
753 TATTGCACAGCTGAATAATCCTTCAGGAAATTAGAAGTCAGAGATGTCA 802  
117 s1LeLysAlaValLeuArgGluProTyrThrValArgValGluAspGln 134  
|||||  
803 CATCAAGGCTGTTTACGGAGCCCTATACAGTCCGTGTGAGAGACCGA 852  
134 ystHMetArgGlyAsnValAlaValPheLysCysIleIleProSerSer 150  
|||||  
853 AAACCATGAGAGCAATGTTCGGTCTTCAAGTGCATTATCCCTCTCG 902  
151 ValGluAlaTyrIleThrValIleSerTrpGluLysAspThrValSerLe 167  
|||||  
903 GTGAGAGCGTACATCACTGCTGCTCATGGAGAAACACACTGTTTACT 952  
167 uValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeu TyrIleL 184  
|||||  
953 TGCTCTCAGATCTAGATTCTTCATCATCATCCAGCGAGCCTTGATATA 1002  
184 ysAspValGlnAsnGluAspGlyLeuTyrAsnTyrArgCysIleThrArg 200  
|||||  
1003 AAGATGTAACAATGAAGATGATGATTAACATACCGCTCATCACCGGG 1052  
201 HisArgTyrThrGlyLysThrArgGlnSerAsnSerAlaArgLeuPheVa 217  
|||||  
1053 CATGATATACACCGAGAGACGACGACGACGACGACGACGACGACGACG 1102  
217 lSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAspHis 234

|||||  
1103 ATCAACACCGAGGAACATCAGCCCATCATCTACTGATGGCTTTACCATC 1152  
234 rGlyAlaMetAlaGlyAlaArgValGluLeuProCysLysAlaLeuGly 250  
|||||  
1153 GCAAGACCATGGCTGGCAGCGCTGTGAGAGTGCCTTGCAAGCCCTCGGG 1202  
251 HisProGluProAspTyrArgTrpLeuLysAspAsnMetProLeuGluLe 267  
|||||  
1203 CACCTGAGCCAGATTACCGCTGCTGAAGAGCAACATGCTCGGAAC 1252  
267 uSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGluAsnIleA 284  
|||||  
1253 TTCAGGAGAGTTCGAAAGACCGTACGCGGCTGCTCATTAGAACAATTC 1302  
284 rGProSerAspSerGlySerTyrValCysGluValSerAsnArgTyrGly 300  
|||||  
1303 GCCCTCGGACTCAGCAGCTATGTTGTGAAGTGTCCACAGATACGGA 1352  
301 ThrAlaLysValIleGlyArgLeuTyrValLysGlnProLeuLysAlaTh 317  
|||||  
1353 ACTGCTAAGGTGATAGCCCGCTGTACTGTAACAGCCACTGAAAGCCAC 1402  
317 lIleSerProAlrGlyValLysSerSerValGlySerGlnValSerLeus 334  
|||||  
1403 CATCAGTCCAGAGAGGTAAAGCAGCGTGGTAGCCCAATCTTCTTGT 1452  
334 erCysSerValThrGlyThrGluAspGlnGluLeuSerTrpTyrArgAsn 350  
|||||  
1453 CCTGAGCGGTGACAGAACTGAGCCAGCAACTCTCTGTGTACCGCAAT 1502  
351 GlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyIleAsnHi 367  
|||||  
1503 GGTGAATCTTCACCTCGAATAAATGTGAGATCACAGGAGATCAACCA 1552  
367 sGluAsnLeuIleMetAspHisMetValLysSerAspGlyGlyAlaTyrG 384  
|||||  
1553 CGAAACCTTATATATGATCATGTCATAAAGTGACGGGGCCCATACC 1602  
384 lncYsPheValArgLysAspLysLeuSerAlaGlnAspTyrValGlnVal 400  
|||||  
1603 AGTGCTTTGTGGCAAGAGCAAGCTGTCCGCTCAAGACTATAGTCAGGTG 1652  
401 ValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGluLysVa 417  
|||||  
1653 GTCTTGGAAGATGAACTCCAAATTTATTTCTGCTTTAGTGAAGAAGT 1702  
417 lValSerProAlaGluProValSerLeuMetCysAsnValLysGlyThrP 434  
|||||  
1703 GGTAGTCCAGCAGACCGCGATTTCCTTATGTGCACGTCGAAGGAACAC 1752  
434 roLeuProThrIleThrTrpThrLeuAspAspAspProIleLeuLysGly 450  
|||||  
1753 CTTTGCCACGATTCAGTGGAACCCGTGAGATGACCCGATTTCTAAGGT 1802  
451 GlySerHisArgIleSerGlnMetIleThrSerGluGluAsnValIse 467  
|||||  
1803 GGCAGTCAACCGATCAGCCAGATGATCAGTCCGAGGGGGAACGTGGTCAG 1852  
467 rTyrLeuAsnIleSerSerSerGlnValArgAspGlyGlyValTyrArgC 484  
|||||  
1853 CTACCTGAACATCTCCACCTCCAGAGTCCGGAGCGGGAGTATACGGCT 1902  
484 ystThrAlaAsnAsnSerAlaGlyValIleLeuTyrGlnAlaArgIleAsn 500  
|||||  
1903 GCACGTGCCAACAACTCGCGGAGTGTCTGTACCAAGGCTCGAATAAAC 1952  
501 ValArgGlyProAlaSerIleArgProMetLysAsnIleThrAlaIleAl 517  
|||||  
1953 GTAAAGGGCCCTGCAAGCATTCGACCAATGAAAAACATCACAGCAATATAC 2002  
517 aGlyArgAspThrTyrIleHisCysArgValIleGlyTyrProTyrTyrS 534  
|||||



17 rgluAspLeuHisSerSerLeuTyrPheValAsnAlaSerLeuGlnIuV 34  
|||||  
503 TGAAGACTTCACTCCAGCCTTACTTGTTCATGTCATCTCTCCAAAGAG 552  
34 a1ValPheAlaSerThrThrGlyThrLeuValProCysProAlaIaGly 50  
|||||  
553 TAGGTTCGCAGCACCAGGGGACTGTGGCTCCCTCCCGCAGCGAGC 602  
51 IleProProValThrLeuArgTrpTyrLeuAlaThrGlyGlnIuIleTy 67  
|||||  
603 ATCCCTCTGTGACTCTCAGTGTGATGCTACCTACCCAGCGGCGAGATCTA 652  
67 AspValProGlyIleArgHisValHisProAsnGlyThrLeuGlnIleP 84  
|||||  
653 CGATGTCCCGGGATCCGCCAGCTCCACCCCAAGCGCATCTCCAAATTT 702  
84 heProheProProSerSerPheSerThrLeuIleHisAspAsnThrTyr 100  
|||||  
703 TCCCTTCCCTCTTCAAGCTTCAGTACCTTAATCCATGATTAATCTTAT 752  
101 TyrCysThrAlaGluAsnProSerGlyLysIleArgSerGlnAspValHi 117  
|||||  
753 TATTTCACAGCTGAATAATCCTTCAGGGAAATTAGAAGTCAGATGTCCA 802  
117 sIleLysAlaValLeuArgGluProTyrThrValArgValGlnuSpGln 134  
|||||  
803 CATCAAGCTGTATTACGGAGCCCTATACAGTCCGTGTGGAGAGCCAGA 852  
134 ystHmeArgGlyAsnValAlaValPheLysCysIleIleProSerSer 150  
|||||  
853 AAACCATGAGAGCAATGTCGGCTTCACAGTCATTAATCCCTCCCTCG 902  
151 ValGluAlaTyrIleThrValValSerTrpGluLysAspThrValSerLe 167  
|||||  
903 GTGGAGGCGTACATCAGTCTCTCATGAGGAGAAAGACACTGTTCCT 952  
167 uValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeuYrIleL 184  
|||||  
953 TGTCTCAGGATCTAGATTCTCATCATCCATCCAGGAGCCTGTATATA 1002  
184 yAspValGlnAsnGluAspGlyLeuTyrAsnTyrArgCysIleThrArg 200  
|||||  
1003 AAGATGTACAGATGAAGATGGATTGTATTACTACCGCTGCATCCAGCG 1052  
201 HisArgTyrThrGlyuThrArgGlnSerAsnSerAlaArgLeuPheVa 217  
|||||  
1053 CATCGATACACCGAGAGAGAGCGAGCAGCAAGCGCCAGACTTTTGT 1102  
217 lSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAspHisA 234  
|||||  
1103 ATCAGACCCAGCACTCAAGCCCATCCATAGTGATGGATTGACCATTC 1152  
234 rglYsAlaMetAlaGlyGlnArgValGluLeuProCysLysAlaLeuGly 250  
|||||  
1153 GCAAAAGCATGGCTGGGCAAGCGGTGTGAGAGTGCCTTGCAAAAGCGCTGG 1202  
251 HisProGluProAspTyrArgTrpLeuLysAspAsnMetProLeuGln 267  
|||||  
1203 CACCCGTGAGCAGATTAACCGCTGCTGAAGACACATGCCCCCTGGAAGT 1252  
267 uSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGluAsnIleA 284  
|||||  
1253 TTGAGGAGAGTTCCAGAAAGCGTGACGGGGCTGCTCATTTGAGAAACATT 1302  
284 rqrProSerAspSerGlySerTyrValCysGluValSerAsnArgTyrGly 300  
|||||  
1303 GCCCCTCGGACTCAGCGAGCTATGTTGTGAAGTGTCACACATACGGA 1352  
301 ThrAlaLysValIleGlyArgLeuTyrValLysGlnProLeuLysAlaTh 317  
|||||  
1353 ACTGCTAAGGTGATAGCGCGCTGTACGTGAACACAGCCACTGAAGGCAC 1402

317 rIleSerProAlaLysValLysSerSerValGlySerGlnValSerLeus 334  
|||||  
1403 CATCAGTCCAGAGAAAGTTTAAAGCAGCGTGGTAGCCCAAGTTCTCTGT 1452  
334 erCysSerValThrGlyThrGluAspGlnIuLeuSerTrpTyrArgAsn 350  
|||||  
1453 CCTGACGGGTGACAGAACTGAGACCGAGCAACTCTCTGCTACCGCAAT 1502  
351 GlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyIleAsnHi 367  
|||||  
1503 GGTGAATCCTCAACCTCGAAGAAATGTGAGATCAACAGGAGATCAACA 1552  
367 sGluAsnLeuIleMetAspHisMetValLysSerAspGlyAlaIArg 384  
|||||  
1553 CGAAACCTTATATGATGCATGCATGCTCAAAAGTGACGGGGCGCATACC 1602  
384 lncSspHeValArgLysAspLysLeuSerAlaGlnAspTyrValGlnVal 400  
|||||  
1603 AGTGCTTTGTGCGCAAGCAAGCTGTCCGCTCAAGACTATGTGCGAGTG 1652  
401 ValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGluLysVa 417  
|||||  
1653 GTCTTGAAAGATGAAGACTCCAAATATTTCTGCTTACGAAAGAGT 1702  
417 lValSerProAlaGluProValSerLeuMetCysAsnValLysGlyThrP 434  
|||||  
1703 GGTGAGTCCAGCAGAGCCGTTTCCCTTATGTGCAACGTGAAGGGAACAC 1752  
434 rLeuLeuProThrIleThrTrpThrLeuAspAspProIleLeuLysGly 450  
|||||  
1753 CTTTGCCAGCAGTACAGTGAACCTGAGGATGACCCGATTTTCAAGGT 1802  
451 GlySerHisArgIleSerGlnMetIleThrSerGluLysAsnValValSe 467  
|||||  
1803 GGCAGTCAACCGCATCAAGCAGATGATCAGTCGAGGAGAAAGTCTCAG 1852  
467 rTyrLeuAsnIleSerSerSerGlnValArgAspGlyGlyValTyrArgC 484  
|||||  
1853 CTACCTGAAATCTCAGCTCCAGCTCCGAGTCCGGAGGGGAGTCTACCGT 1902  
484 ystHraIaAsnAsnSerAlaGlyValValLeuTyrGlnAlaArgIleAsn 500  
|||||  
1903 GCATGCGCAACACTCGCGGAGTGTCTCTGTACCAAGCTCGAATTAAC 1952  
501 ValArgGlyProAlaSerIleArgProMetLysAsnIleThrAlaIleAl 517  
|||||  
1953 GTAGAGGCGCTCGAAGCATTCAGCAATGAATAACATCACACCAATATAC 2002  
517 acGlyArgAspThrTyrIleHisValArgValIleGlyTyrProTyrTyrS 534  
|||||  
2003 AGGAGGGGACATACATTACTGTCTGTGATTGGCTATCCGTATTACT 2052  
534 erIleLysTrpTyrLysAsnSerAsnLeuLeuProPheAsnHisArgGln 550  
|||||  
2053 CCATTAAATGGTACAAAGACTTAACCTGCTCTTCAACCAAGCCGCCAA 2102  
551 ValAlaPheGluAsnAsnGlyThrLeuLysLeuSerAspValGlnLysG 567  
|||||  
2103 GTGCATTTGAAACAATGGAACCTTAAACTTTCAGATGTGCAAAAGGA 2152  
567 uValAspGluGlyLysIleTyrThrCysAsnValLeuValGlnProGlnLeuS 584  
|||||  
2153 AGTGACGAGGGGAGATACAGCTGCACAGTGTGTTCAACCAACAACTCT 2202  
584 erThrSerGlnSerValHisValThrValLysValProProPheIleGln 600  
|||||  
2203 CCACAGCAGCAGAGCGTCCACCTGACCGTGAAGATTCGCCCTTTCATACAA 2252  
601 ProPheGluProProArgPheSerIleGlyGlnArgValPheIleProCy 617  
|||||  
2253 CCCTTGAGTTTCCAGATTTCTCATTTGGGAGCGGCTTTCATCCCTCG 2302  
617 sValValValSerGlyAspLeuProIleThrIleThrTrpGlnLysAspG 634



149 CCTGAGATGTTGGACACGCTCTACTTTGTAATGACTCTTGACG 198  
33 luvAlvalPhealaserThrThrGlyThrLeuValProCysProAlaAla 49  
AGGTACCTTTTCCAGCTCCGTGGGGGTGTGTGCTCCGCCGGCGG 248  
50 GlyIleProValThrLeuArgTrpTyrLeuAlaThrGlyGluIle 66  
GGCTCTCCCGACGGCGCTTCGATGTACTGGCCACAGGGGACGACAT 298  
66 eTyrAspValProGlyIleArgHisValHisProAsnGlyThrLeuGln 83  
CTACAGACGTCCCGACATCCGGACGTCCAGCCACAGGACGCTGACG 348  
83 LephProPheProPheSerPheSerThrLeuIleHisAspAsnThr 99  
TCTACCCCTCTCCCTCCGCTTCAATACCTTATCCACGACAATGAC 398  
100 TyrTyrCysThrAlaGluAsnProSerGlyLysIleArgSerGlnAspVal 116  
TACTTCTGACCGCGGACAGACGCTCCGGCAAGATCCGAGC TCCACAT 448  
116 HisIleLysAlaValLeuArgGluProTyrThrValArgValGluAspG 133  
CCGCGTCAAGACAGTTTTCAGGGAACCTACACCGCTCCGGGTGAGGATC 498  
133 InLysThrMetArgLysAsnValAlaValPheLysCysIleIleProSer 149  
AAAGGTCAATCGTGGCAACGTGGCGCTTCAAGTGTCTATCCCTCT 548  
150 SerValGluAlaTyrIleThrValValSerTrpGluLysAspThrVal 166  
TCAGTGCAGGAATATGTAGCGTTGTATCTGGGAGAAAGACACAGCTC 598  
549 ThrValSerGlySerArgPheLeuIleThrSerThrGlyAlaLeuTyrI 183  
CATCATCCACAGAAACAGCTTTTATTACTTACCACGCGCGCTGTACA 648  
183 lElasAspValGlnAsnGluAspGlyLeuTyrAsnTyrArgCysIleThr 199  
TCTCTGACGTACAGAGAGAGAGCGCTTCCACCTATCGCTGATCACC 698  
200 ArgHisArgTyrThrGlyGluThrArgGlnSerAsnSerAlaArgLeuPh 216  
AAGCAGCATATAGCGGGAGACCCCGGACAGCATAGGGGACCGCTCTC 748  
699 eValSerAspProAlaAsnSerAlaProSerIleLeuAspGlyPheAspH 233  
TGTGACAGACCTGCTGAGTCCGATCCACCATCTCGATGGTTCCACT 798  
749 lAsArgLysAlaMetAlaGlyGlnArgValGluLeuProCysLysAlaLeu 249  
CCCAGAGAGTGGGCGCGGACACCGGTGAGCTGCCCTGACCGCGCTCG 848  
799 GlyHisProGluProAspTyrArgTrpLeuLysAspAsnMetProLeuG 266  
GGCTACCTATCCCGGACATCCGCTGCTCAAGATGGCGCGCGCTCCG 898  
849 uLeuSerGlyArgPheGlnLysThrValThrGlyLeuLeuIleGluAsnI 283  
GGCTGACAGCCGCTGACCAAGCCATCACAGGCTGACATCAGAGACT 948  
899 lAsArgProSerAspSerGlySerTyrValLysGluValSerAspArgTyr 299  
TGGGACCGAGAGACGCGCACCTACATTGTGAGTCAACCAACACTTC 998  
949 GlyThrAlaLysValIleGlyArgLeuTyrValLysGlnProLeuLysAl 316  
GGTTGGCAGAGGCCACAGCATCTCATGTCTATGATCCCTTCATGT 1048  
999 aThrIleSerProArgLysValLysSerSerValGlySerGlnValSerL 333  
1049 GACCTGACACCAAGAGAGTGAAGACCGGCTTGGCAGCAGCTCATCC 1098

333 euSerCysSerValThrGlyThrGluAspGlnGluLeuSerTrpTyrArg 349  
TCTCCGTGTCCTGTGACGGGCTCCCGAGATTCCACCATCGGTGTATGC 1148  
1099 AsnGlyGluIleLeuAsnProGlyLysAsnValArgIleThrGlyLeuAs 366  
AACACGAGACTGTGCTGCTGTACAGAGGTCATCTTCATCCGTGGCTAG 1198  
1149 nHisGluAsnLeuIleMetAspHisMetValLysSerAspGlyAlaT 383  
CAACGAGAGCGTGTATCATCTGCGCCAGAGAGCCATTCCGGGGCT 1248  
1199 yGlnCysPheValArgLysAspLysLeuSerAlaGlnAspTyrValGln 399  
ACCAAGTCTGCTTACCCGCAAGGCCACAGCCGCCAGACTTTCCTCATC 1298  
1249 ValValLeuGluAspGlyThrProLysIleIleSerAlaPheSerGly 416  
ATTGCATGTGAGATGGACAGCGCCGACATCGTCTGCTTCAGGAGAA 1348  
1299 sValValSerProAlaGluProValSerLeuMetCysAsnValLysGlyT 433  
GGTGTCAACCCCGGGGAGACAGTCTCATGTGTCGCGCAAGGCGG 1398  
1349 hrProLeuProThrIleThrTrpThrLeuAspAspProIleLeuLys 449  
CCCCGCCGCCCGGTCACCTGCGCTCGACGATGACCCCATCTGGG 1448  
1399 450 GlyGlySerHisArgIleSerGlnMetIleThrSerGlyGly 463  
GATGGAGCGCACCGCACCAACAGTACACCATGTGCGACGGG 1490  
1449

seq\_name: /cgn2\_6/p/tdata/1/ina/PCTUS9\_COMB.seq:PCT-US94-05277-1

seq\_documentation\_block:  
Sequence 1, Application PC/TUS9405277  
GENERAL INFORMATION:  
APPLICANT: Bruskin, Arthur  
APPLICANT: Jarosz, David E.  
APPLICANT: Johnson, Kate  
APPLICANT: Kinzler, Kenneth W.  
APPLICANT: Vogelstein, Bert  
APPLICANT: Zaretsky, James R.  
TITLE OF INVENTION: Antibodies Specific for DCC Gene Product  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05277  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42709  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508 9100  
TELEFAX: 202 508 9299  
TELEX: 197430 BBWB UT  
INFORMATION FOR SEO ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4608 base pairs

OM of: US-08-956-991-11 to: EST: \* out\_format : pfs  
 Date: Apr 25, 2000 9:52 PM

About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=f1amer-p2n.model -DEV=xlp  
 -O=cpn2.1/USPTO.spool/US08956991/runat.25042000.12.5.7.9550/app-query.fasta.1  
 -DB=EST -QPM=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
 -MIMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
 -DELETE=7.000 -START=1 -MATRIX=blsnum62 -TRANS=human40.cdi  
 -LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -ALIGN=15 -MODE=LOCAL  
 -OUTFM=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08956991  
 -NCPU=6 -ICPU=3 -NO\_ALIGN=wait -THREADS=1

## Search information block:

Query: US-08-956-991-11  
 Query length: 1571  
 Database: EST: \*  
 Database sequences: 4538634  
 Database length: 1887831982  
 Search time (sec): 2119.170000

## score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_est27:AI454704	723.00	1456.14	2.3e-72	458	1	AI454704 UI-R-BT0-qk-h-03-0-UI
gb_est1:AI5159035	665.00	1338.23	9.0e-66	427	1	AM159035 za50e05.x1 Xenopus EST
gb_est18:AO003472	604.00	1215.45	6.2e-59	355	1	AJ003472 AJ003472 Selected chr2
gb_gss1:FR0020941	576.00	1150.03	2.7e-55	619	1	286724 F. rubripes GSS sequence
gb_est13:AI106884	411.00	810.29	2.3e-36	646	1	AI106884 GH06134.5prime GH Dros
gb_est18:AM048129	408.50	810.29	1.9e-36	402	1	AM048129 UI-R-BH1-alc-04-0-UI
gb_est18:AA697364	408.50	802.87	6.3e-36	786	1	AA697364 HL02328.5prime HL Dros
gb_est27:AI404957	355.50	698.02	4.1e-30	569	1	AI404957 GH26371.5prime GH Dros
gb_gss1:FR0021997	346.50	678.53	5.0e-29	610	1	AI404868 F. rubripes GSS sequenc
gb_est28:AI542894	345.50	674.61	8.2e-29	696	1	AI542894 SD09407.5prime SD Dros
gb_est17:AM014717	344.00	675.92	4.6e-29	409	1	AM014717 UI-R-BH0-aag-a-07-0-UI
gb_est12:AA325603	332.50	659.93	5.4e-28	298	1	AA325603 EST8915 Cerebellum II
gb_gss1:FR0030733	329.50	648.17	2.4e-27	442	1	AL027102 Fugu rubripes GSS sequ
gb_gss1:FR002934	328.00	645.25	3.1e-27	437	1	286717 F. rubripes GSS sequenc
gb_est28:AI541596	317.50	618.36	1.1e-25	635	1	AI541596 SD02772.5prime SD Dros
gb_gss1:FR0021994	298.50	579.99	1.5e-23	605	1	AI404865 F. rubripes GSS sequenc
gb_gss1:FR0030753	293.50	569.51	5.9e-23	614	1	AL027122 Fugu rubripes GSS sequ
gb_est18:AI498283	289.50	552.72	9.6e-23	450	1	AI498283 tm8h11.x1 NCI_CGAP_BI
gb_gss1:AI406159	285.00	552.72	5.1e-22	585	1	AI406159 GH26371.5prime GH Dros
gb_gss1:FR0029758	283.50	549.16	8.0e-22	605	1	AL027127 Fugu rubripes GSS sequ
gb_gss1:FR002038	275.50	532.79	6.5e-21	602	1	AL014909 F. rubripes GSS sequenc
gb_gss1:CNS010PH	274.00	525.94	1.6e-20	785	1	AL092623 Drosophila melanogast
gb_est13:AI107879	272.00	525.96	1.6e-20	587	1	AI107879 GH05755.5prime GH Dros
gb_est18:AI519340	250.00	480.96	5.0e-18	578	1	AI519340 LD38885.5prime LD Dros
gb_gss1:FR0022007	246.50	477.54	7.8e-18	443	1	AL014678 F. rubripes GSS sequenc
gb_est28:AI543452	246.00	472.94	1.4e-17	570	1	AI543452 SD10207.5prime SD Dros
gb_est1:AI524594	240.00	470.59	1.9e-17	282	1	AV264594 AV264594 RIKEN full.1e
gb_gss1:CNS00058	231.50	433.79	2.1e-15	1101	1	AL050788 Drosophila melanogast
gb_est20:AA812746	226.50	427.76	4.6e-15	116	1	AA812746 LD22445.5prime LD Dros
gb_gss1:CNS02012	224.50	428.80	4.1e-15	549	1	AI014833 F. rubripes GSS sequenc
gb_gss1:CNS01040	224.00	419.78	1.3e-14	997	1	AL098490 Drosophila melanogast
gb_est28:AI531274	222.50	422.79	8.8e-15	649	1	AI531274 SD02242.5prime SD Dros
gb_gss1:HSWC18B12	221.00	440.79	8.7e-16	147	1	X88325 H. sapiens DNA for trapp
gb_gss1:FR0009351	221.00	433.95	2.1e-15	238	1	AL000636 F. rubripes GSS sequenc
gb_est14:AM213058	220.00	420.48	1.2e-14	532	1	AM213058 un98G01.Y1 NCI_CGAP_M
gb_gss1:CNS012Y3	215.50	401.81	1.3e-13	1033	1	AI013165 Drosophila melanogast
gb_est28:AI513189	215.00	408.19	5.7e-14	613	1	AI513189 LD45391.5prime LD Dros
gb_est26:AU004083	215.00	408.08	7.5e-14	711	1	AU004083 AU004083 Bombyx mori F
gb_gss1:FR0022080	214.50	406.85	5.2e-14	544	1	AI014951 F. rubripes GSS sequenc
gb_est12:AI722009	214.00	406.46	7.0e-14	595	1	AI722009 fd18C09.Y1 Sugano Kawa
gb_gss1:CNS0070A	208.00	386.45	9.2e-13	1028	1	AI067489 Drosophila melanogast
gb_gss1:AO815550	207.00	395.52	2.9e-13	470	1	AO815550 HS_5346_A2_B08_SBE6 RH

gb\_gss1:FR0030755 + 206.50 393.93 3.5e-13 489 1 AL027124 Fugu rubripes GSS S  
 gb\_est28:AI515655 + 206.50 393.64 3.7e-13 499 1 AI515659 LD2472.5prime LD D  
 gb\_est38:AL119288 + 203.50 384.88 1.1e-12 599 1 AL119288 DXF2P61A081\_r1 761

seq\_name: gb\_est27:AI454704

## seq\_documentation\_block:

LOCUS AI454704 458 bp mRNA EST 05-JUL-1999  
 DEFINITION UI-R-BT0-qk-h-03-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone  
 UI-R-BT0-qk-h-03-0-UI 3', mRNA sequence.

ACCESSION AI454704

VERSION AI454704.1 GI:4295587

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Scurionathia; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT On Mar 20, 1998 this sequence version replaced g1:2980496.  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel.: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Oligo- or track not found. Not a site shown in beginning of sequence  
 is likely internal to the message. cDNA library preparation: M.  
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available  
 through Research Genetics (www.resgen.com) This clone is also  
 available through the I.M.A.G.E. Consortium at LNL  
 (info@image.llnl.gov). IMAGE ID-I788947  
 Seq primer: M13 forward  
 POLYA-No.

## FEATURES

## SOURCE

1..458 Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-BT0-qk-h-03-0-UI"

/cclone\_lib="UI-R-BT0"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site\_1: Not I; Site\_2: Eco RI; This library

(UI-R-BT0) consists of a mixture of individually tagged

normalized libraries constructed from rat hippocampus,

thalamus, mid-brain, medulla, corpus striatum, cerebral

cortex and testis. The tag used to identify the source

tissue is a string of 3-6 nucleotides present between the

Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture. This library was then subtracted using a

driver consisting of a mixture of all clones from UI-R-A0,

UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and

UI-R-C2p."

ORIGIN

BASE COUNT 107 a 112 c 123 g 116 t

## alignment\_scores:

Quality: 723.00

Ratio: 5.021

Percent Similarity: 94.737

Percent Identity: 91.447

alignment\_block:

US-08-956-991-11 x AI454704/rev ..

Align seg 1/1 to reverse of: A1454704 from: 1 to: 458

```

525 CysArgValIleGlyTyrProTyrTyrSerIleTyrTrpTyrLysAsnSe 541
541 rAsnLeuLeuProPheAsnHisArgGlnValAlaPheGluAsnAsnGlyT 558
457 TGCTGTGATTTGGTTACCCCTATTACTCCATCAAGTGTACAAAGACGA 408
407 TAACCTGCTTCTTTCAACCCAGCAGAGTGGCTTTGAGAACATATGTA 358
558 hLeuLysLeuSerAspValGlnLysGluValAspGluGlyTyrThr 574
357 CTCCTGAACACTCAGATGTGCAGAAAGAGTTGACGAGGAGATACACA 308
575 CysAsnValLeuValGlnProGlnLeuSerThrSerGlnSerValHisVa 591
307 TGTACGCTGCTGTGACGCCACACACTCTCCACACGACGAGAGTCCAGCT 258
591 lThrValLysValProProPheIleGlnProPheGluProArgPheS 608
257 GACAGTCAAAGTTGCTCTTTATCCAACTTTGAGTTCCGACATTTT 208
608 eTlIGLyGlnArgValPheIleProGlyValValValSerCysAsnLeu 624
207 ATATGGGTACGGGGATTCATCCATGTGTGTGTGTGTGTGTGTGTGT 158
625 ProIleThrIleThrTrpGlnLysAspGlyArgProIleProGlySerLe 641
157 CCCATCACCATCAGTGGCAGAGAGATGCCCTCGATCCGACGACAGCTT 108
641 uGlyValThrIleAspAsnIleAspPheThrSerSerLeuArgIleSera 658
107 CGGTATACCATGCAGAACATCGACTTCACGACGCTCTTGAGATCTCA 58
658 sLeuSerLeuMetHisAsnGlyAsnValTyrTyrCysIleAlaArgAsnG 674
57 ACCTCTCACTGATGCACAACGGGATATACATGATGATGAGCAGAG 8
675 AlaAla 676
7 GCGGCC 2
seq_name: gb_est41:AM159035
seq_documentation_block:
LOCUS AM159035 427 bp mRNA EST 05-NOV-1999
DEFINITION z50605.x1 Xenopus EST library Xenopus laevis cDNA clone z50605
5' mRNA sequence.
ACCESSION AM159035
VERSION AM159035.1 GI:6271064
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;
Xenopus.
1 (bases 1 to 427)
Schultz,K., de la Bastide,M., Huang,E.N., Nascimento,L., Preston,R.,
Shah,R., Staby,I., Shekher,M., Spiegel,L., Vil,M.D. and
McCombie,M.R.
Expressed sequence tags from Xenopus
unpublished (1999)
On May 18, 1998 this sequence version replaced gi:3137856.
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cschl.org
Plate: z50 row: e column: 05
Seq primer: M13 universal forward primer
High quality sequence stop: 427.

```

```

FEATURES
    source
        Location/Qualifiers
            1..427
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone="z50605"
                /clone_lib="Xenopus EST library"
                /tissue_type="total brain tissue"
                /cell_line="W22-TGA"
                /dev_stage="tadpole"
                /note="vector: Lambda Zap II; site_1: XbaI; This library
                was supplied by Holly Cline (Cold Spring Harbor Labs).
                cDNA synthesis with oligo dt Xba I (Xba I cloning site).
                RNA: stage 50-56 tadpoles, total brain tissue, GTC
                extraction method."
BASE COUNT      114 a      122 c      91 g      100 t
ORIGIN
alignment_scores:
    Quality: 665.00      Length: 139
    Ratio: 4.890      Gaps: 1
    Percent Similarity: 97.842      Percent Identity: 92.806
alignment_block:
US-08-956-991-11 x AM159035
Align seg 1/1 to: AM159035 from: 1 to: 427
410 lIleSerAlaPheSerGluLysValValSerProAlaGluProValSerLe 426
17 ATCTCGCATTTAGCGAGAAAGGTGTGACCCCGGAGAACCTGTCTCCT 66
426 uMetCysAsnValLysGlyThrProLeuProThrIleThrTrpThrLeua 443
67 TATGTCAACGTMAAGAACGCTCTTCCGACGATCACTTGAGAGCTGG 116
443 sPasAspProIleLeuLysGlyLysSerHisArgIleSerGlnMetIle 459
117 ACGACAGCCCATTTATCAGAACGCCGACACACAGATAAGTAGTTAATC 166
460 ThrSerGluGlyAsnValValSerTyrLeuAsnIleSerSerSerGlnVa 476
167 ACCTTGAGGGCAACGTTGTCACTTACCTGACATCACCACATCACTCAAGT 216
476 lArgAspGlyGlyValTyrArgCysThrAlaAsnAsnSerAlaGlyVal 493
217 CCAGATGGCGGGGTGTACCATGCACCGCTTAACAACCTGTGGGCGCTCG 266
493 aLLeuTyrGlnAlaArgIleAsnValArgGlyProAlaSerIleArgPro 509
267 TCTTGTACAGGCTCGAATAAAGT.....CTTCAAGTATCCGCTCCA 309
510 MetLysAsnIleThrAlaIleAlaGlyArgAspThrTyrIleHisCysAr 526
310 ATGAAACATCATCGCGCATAGCTGGGGGGGACACATCATCACTGCGG 359
526 gValIleGlyTyrProTyrTyrSerIleLysTrpTyrLysAsnSerAsnL 543
360 CGTCATGTGATATCCCTACTACTCATCAAGGTGACAAAACCTTAACG 409
543 eLeuLeuProPheAsnHis 548
410 TCCTCCCATTTAATCAC 426
seq_name: gb_est18:AJ003472
seq_documentation_block:
LOCUS AJ003472 355 bp mRNA EST 01-APR-1998
DEFINITION AJ003472 selected chromosome 21 cDNA library Homo sapiens cDNA
clone MP112-218, mRNA sequence.
ACCESSION AJ003472
VERSION AJ003472.1 GI:2769503
KEYWORDS EST.

```